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　５ 年　 ６ 月 ２２ 日現在

機関番号：１２６０１
研究種目：基盤研究(C)（一般）
研究期間：2018～2022
課題番号：１８Ｋ１１６００

研究課題名（和文）連続空間ゲームにおける深層学習を利用した強化学習

研究課題名（英文）Reinforcement Learning Using Deep Learning in Continuous Space Games

研究代表者
田中　哲朗（Tanaka, Tetsuro）

東京大学・情報基盤センター・准教授

研究者番号：６０２５１３６０

交付決定額（研究期間全体）：（直接経費）　　2,300,000円

研究成果の概要（和文）：デジタルカーリングを用いた研究の基礎として，カーリングの不確実性を排除した「決定的なデジタルカーリング」を提案し，そのゲームの勝敗に関する有益な知見を得たこと，不完全情報ゲームを扱うための階層型強化学習の有効性を検証するために，麻雀を用いた階層型強化学習の評価を行い，Optunaのようなハイパーパラメータ自動最適化フレームワークの有効性を確認したこと，そしてGANを用いたタワーディフェンスゲームの自動生成において有効性を検証したこと，いくつかの不完全情報ゲームのナッシュ均衡戦略をもとめたことなどが挙げられる．これらの研究成果はプログラムが公開され，今後の研究者に利用可能となっている．

研究成果の学術的意義や社会的意義
本来の研究目的である連続空間ゲームにおける深層学習を利用した強化学習における有効な学習手法の提案は実現できなかったため，学術的には大きな成果をあげることはできなかったといえる．一方で，社会的意義としては，連続空間ゲームであるカーリングの性質を考察することにより，学習アルゴリズムにおいて考慮すべき点などを指摘した点，連続空間ゲームと深い関連を持つ，不完全情報ゲームのいくつかについて，強解決をおこなったり，ナッシュ均衡戦略を求め，その解析結果を公開することにより，それらのゲームを題材に深層学習を利用した強化学習をおこなう際の評価の指標となる「正解」を与えた点など，一定の成果を果たした．

研究成果の概要（英文）：As a foundation for research using digital curling, we proposed "deterministic digital curling," which eliminates the uncertainty of curling and obtained valuable insights into the game's outcomes. To validate the effectiveness of hierarchical reinforcement learning for handling incomplete information games, we conducted an evaluation using Mahjong and confirmed the effectiveness of hyperparameter automatic optimization frameworks like Optuna. Additionally, we verified the effectiveness of using GANs for the automatic generation of tower defense games and identified Nash equilibrium strategies for several incomplete information games. These research achievements have been made publicly available through published programs, making them accessible to future researchers.

研究分野：ゲーム情報学

キーワード：連続空間ゲーム　強化学習　不完全情報ゲーム　ナッシュ均衡　強解決

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

囲碁，将棋，チェスなどの離散で有限な状態空間，アクション空間を持つゲームに関して，畳み込みニューラルネットワーク(C NN)を用いたポリシーネットワーク，バリューネットワークを構築し，モンテカルロ木探索を用いた自己対戦結果を用いた強化学習で学習させる手法が成功をおさめていた．

また伝統的なボードゲーム以外にも，深層強化学習 DQN (Deep Q-Learning) が初期のテレビゲームである ATARI ゲームを評価対象にしたように，適用可能なゲームの範囲を広げる研究が行われていた．そのような中で，カーリングをモデル化したデジタルカーリングという対象が考案されて，定期的に競技がおこなわれるようになってきた．

デジタルカーリングにおいては，機械学習を用いた手法や，探索を用いた手法が成功してきたが，深層強化学習の適用はカーリングのような連続状態空間，連続アクション空間を対象にした場合は容易ではないと考えられていた．また，カーリングの持つ不確実性を適切に扱うための手法も必要になると考えられていた．

２．研究の目的

カーリングのような連続状態空間，連続アクション空間を対象にしたゲームについて，効果的な畳み込みニューラルネットワーク(CNN)の構成法を構築し，連続アクション空間でも可能な探索手法を提案，評価することが，研究目的である．

一般に連続空間の強化学習タスクとして用いられる倒立振子などと比較すると，入力の次元が大きく，2 人ゲームであり，またアクション空間における累積報酬の極大点が複数存在する性質を持つカーリングのようなタスクをどう扱うかという問題の難しさがある．

３．研究の方法

カーリングは連続状態空間，連続アクション空間を対象にしたタスクの中でも，不確実性も含むため，まずはカーリングから不確実性を取り除いた時に，どのような性質があるかどうかを調べる．

また，GPU 搭載並列計算機を用いて，不完全情報性を持つゲーム，不確実性を持つゲームについて，従来の深層強化学習手法の評価をおこなうこととした．

最善の戦略が解けているゲームを対象に強化学習をおこなうのは，実用的な意味は大きくないが，従来手法との比較という相対的な評価だけでなく，最善戦略に対してどこまで達成できているかという絶対的な評価をおこなうことができる．

４．研究成果

本来の研究目的である連続空間ゲームにおける深層学習を利用した強化学習における有効な学習手法の提案は実現できなかったため，本来の目的については成果をあげることはできなかったといえる．

一方で,「デジタルカーリング」のゲームとしての性質を調べるために，ショットの誤差を 0 にした「決定的なデジタルカーリング」を提案して，そのゲームの勝敗に関して,「フリーガードゾーンのないルールでは第 1 エンドの後攻チームが勝てる」,「あるエンドで先攻チームが後攻チームに 2 点以上取らせない戦略が存在する」などの部分的な結果を得たことは，デジタルカーリングの難しさを考える上では貴重な知見を得たと言える.

また,「十六むさし」や「量子アンパンマンのはじめてしょうぎ」などのいくつかのゲームの強解決をおこない，状態数が 537,103 個の R-Rivals というランダムな要素を含まない同時手 番ゲームの 2 人零和ゲームのナッシュ均衡戦略を求めた.これらの研究成果はプログラムや解析結果のデータを公開することによって，今後の研究者がゲームを対象にした強化学習の評価をおこなう際の題材として利用可能にしている.

他には，グリッド世界や，麻雀類似ゲームを対象にした深層強化学習の評価をおこない，適切なバリュー関数を求めるために，Optuna のようなハイパーパラ メータ自動最適化フレームワークを利用することが有効であるという知見を得た.

これらの研究成果のプログラムを公開することによって，今後の研究者が容易に利用できる点は社会的意義があるといえる.

５．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計8件（うち招待講演　0件／うち国際学会　0件）

１．発表者名 田中哲朗
２．発表標題 R-Rivals のナッシュ均衡戦略
３．学会等名 第27回ゲームプログラミングワークショップ 2021
４．発表年 2021年

１．発表者名 Yueming Xu, Tetsuro Tanaka
２．発表標題 Procedural Content Generation for Tower Defense Games:a Preliminary Experiment with Reinforcement Learning
３．学会等名 第27回ゲームプログラミングワークショップ 2021
４．発表年 2021年

１．発表者名 清水大志，田中哲朗
２．発表標題 深層強化学習を用いた麻雀プレイヤの構築
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 2020
４．発表年 2020年

１．発表者名 田中哲朗
２．発表標題 量子「アンパンマンのはじめてしょうぎ」の強解決
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 2020
４．発表年 2020年

１．発表者名
田中哲朗

２．発表標題
十六むさしの強解決

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 2020

４．発表年
2020年

１．発表者名
清水 大志 , 田中 哲朗

２．発表標題
麻雀のポリシー関数に適したネットワークモデルの構築と評価

３．学会等名
情報処理学会ゲームプログラミングワークショップ2019

４．発表年
2019年

１．発表者名
高岡 峻 , 田中 哲朗

２．発表標題
グリッド世界を用いた階層型強化学習の評価

３．学会等名
情報処理学会ゲームプログラミングワークショップ2019

４．発表年
2019年

１．発表者名
田中哲朗

２．発表標題
決定的なデジタルカーリングの戦略

３．学会等名
カーリング科学ワークショップ

４．発表年
2018年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

r-rivals検証コード
https://github.com/tanakat01/r-rivals
すずめ雀強化学習実験プログラム
https://github.com/minnsou/suzume-jong
量子「アンパンマンのはじめてしょうぎ」の後退解析プログラム
https://github.com/tanakat01/quantum_anpanman
十六むさし後退解析プログラム
https://github.com/tanakat01/16musashi
十六むさし局面検索
https://gps.tanaka.ecc.u-tokyo.ac.jp/16musashi/
ミニ麻雀環境
https://github.com/u-tokyo-gps-tanaka-lab/mini_mahjong
「グリッド世界を用いた階層型強化学習の評価」 実験コード
https://github.com/u-tokyo-gps-tanaka-lab/gridworld_for_HRL

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