

# ぷよぷよテトリスにおけるゲーム AI アルゴリズムの比較

岩井 駿人† 西川 瑳亮‡ 佐々木 雄大‡ 佐藤 健哉‡

†同志社大学工学部情報システムデザイン学科

‡同志社大学大学院理工学研究科情報工学専攻

## 1 はじめに

近年の人工知能 (AI) の進化によって、様々なビデオゲームにおいて人間のプレイヤーよりも好成績を出すゲーム AI が開発され、その一例として、Zetris や Cold Clear などのテトリスのゲーム AI が挙げられる。これらのテトリスのゲーム AI はテトリスのプロを圧倒的に凌駕する強さを実現しており、ぷよぷよテトリスというゲームを起点にその実力が注目されるようになった。ぷよぷよテトリスは、ぷよぷよとテトリスの二つのパズルゲームを組み合わせたゲームで、プレイヤーはそれぞれ得意なゲームを選択して対戦することができる。

ぷよぷよに対してパズルゲームで有用性が示されているアルゴリズムを適用した例は存在する。例えば、2手先までの全探索アルゴリズムを実装している mayah[1] が従来手法として挙げられる。しかし、それらのアルゴリズム同士の比較やぷよぷよテトリスにおける有用性を示した研究はない。ぷよぷよとぷよぷよテトリスは戦術が大きく異なり、ゲーム AI のアルゴリズムはそれぞれのゲームに最適化されると考える。そこで本研究では、ゲーム AI で扱われる複数のアルゴリズムをテトリス相手のぷよぷよに適用し、アルゴリズムの有用性を比較することを目的とする。

## 2 関連研究

### 2.1 モンテカルロ木探索を用いたぷよぷよ AI

添島ら [2] は、ぷよぷよに木探索アルゴリズムの1種であるモンテカルロ木探索を適用した。この研究の結果から、モンテカルロ木探索に優位性が見られなかったことが示されている。しかし、テトリス相手におけるぷよぷよの対戦では、連鎖構築能力だけでなく状況判断能力が必要とされ、この研究とは異なった結果が

得られる可能性がある。

### 2.2 深層強化学習によるぷよぷよ AI

高田ら [3] は、AlphaZero のアプローチである深層強化学習を用いて、ぷよぷよの連鎖構築に特化したゲーム AI を作成した。連鎖構築能力の向上が見られた反面、学習データの不足によって十分に成長できず、従来手法を上回ることができなかったことが報告されている。

## 3 実験

### 3.1 概要

テトリス相手の試合におけるぷよぷよ側は、自フィールドの状態や次に来るぷよの情報のほかに、テトリス側の攻撃などが意思決定の要因となる。本研究では、自フィールドの状態及び次に来るぷよ、また相手の攻撃の量の3つを、ぷよぷよ側の取る行動の要因となる入力とする。そしてこれらの入力から、最善の行動となる手を選択する必要がある。

ぷよぷよは3手目以降にランダム要素がある不確定ゲームに分類される。また、連鎖の処理が複雑であるため、一般的な探索アルゴリズムが全て有用であるとは言えないため、その中で有用であると推測されるアルゴリズムは限られる。本研究では、後述するアルゴリズム6つを取り上げ比較を行う。

### 3.2 検討するアルゴリズム

前述の mayah では、2手先の全探索の際にテンプレートマッチングなども行い評価を行っている。しかし本研究では2手先の全探索を行い、得られたスコアやぷよ同士の連結の数を評価する関数を用いて、mayah のような探索方法を実装したものを従来手法とし、それを1つ目のアルゴリズムとする。

次に2つ目が、対戦データを使った教師あり学習である。ぷよぷよテトリスの対戦の記録が YouTube 上に動画データとして多々存在している。それらを扱える形に変換し10000手分以上の対戦データを用意し、学習を行った。

3つ目に検討するアルゴリズムは、教師データを事

### Comparison of Game AI Algorithms in Puyo-Puyo Tetris

Hayato Iwai†, Sasuke Nishikawa†, Yudai Sasaki‡ and Kenya Sato‡

†Faculty of Science and Engineering, Doshisha University

‡Graduate School of Science and Engineering, Doshisha University

表 1: アルゴリズムごとのスコア・生存ターン・最大連鎖

評価項目\アルゴリズム	従来手法	教師あり	強化学習	MCTS	MCTS × 教師あり	MCTS × 強化学習
スコア	663	474	563	3273	3393	2361
生存ターン	24.99	27.28	29.68	50.53	51.36	42.61
最大連鎖	1.11	1.17	1.20	2.2	2.21	2.03

前に用意しない機械学習の手法である強化学習である。対戦のシミュレーションを繰り返し、自らを強化するように学習を行う方法である。本研究では、DQN という強化学習の手法を用いて学習を行った。

そして4つ目が、モンテカルロ木探索 (MCTS) である。モンテカルロ木探索とは、モンテカルロ法を使った探索アルゴリズムの1種であり、AlphaGo[4] などのアルゴリズムでも用いられているものである。この探索では、ランダムに手を進めた場合の結果を活用し探索を行うため、数手先まで見据えた手を選択することができる。

最後に5つ目と6つ目は、モンテカルロ木探索と機械学習を併用する手法である。モンテカルロ木探索では、前述の通りランダムに手を進めることによって手の評価を行う。しかし、実際に操作を行う際にはぶよを置く位置と向きを考慮した22通りがあり、その候補からランダムに手を打つと効率が悪くなる場合がある。そこで、AlphaGoのアプローチのような、ランダムではなく機械学習によって得られた評価関数を用いる手法が存在する。本研究では、5つ目をモンテカルロ木探索と教師あり学習を用いる手法と、6つ目をモンテカルロ木探索と強化学習を用いる手法とする。

### 3.3 ニューラルネットワークのモデル

本研究で扱う機械学習では、以下の表2のようなニューラルネットワークの構成を使用する。今回の実験ではぶよぶよのフィールドサイズを縦12マス×横6マスとし、ぶよは4色とする。そのためゲームの状態を、おじゃまぶよや相手の攻撃なども含めた23種類のチャネルに変換し、12×6×23の3次元配列として入力とする。

表 2: ニューラルネットワークの構成

入力 [12,6,23]
畳み込み [12,6,128]
最大値プーリング [6,3,128]
畳み込み [6,3,64]
最大値プーリング [3,1,64]
全結合 [22]
着手確率

## 4 評価

### 4.1 評価方法

本研究では、対戦のシミュレーションをプログラム上で評価を行った。シミュレーションでは、事前に用意した攻撃の記録を使用し、実際の対戦を模倣する。そして、100試合の対戦シミュレーションで評価を行った。評価の基準となるのは、シミュレーションを生き残ったターン数、発生したスコアと連鎖とする。すべてのアルゴリズムでそれぞれシミュレーションを行い、生き残ったターン数から生存能力、スコアや連鎖から連鎖構築能力などそれぞれのアルゴリズムの特徴を比較する。

### 4.2 評価結果

評価結果の一部を表1に示す。結果から、ぶよぶよのように2手先以降が不完全な情報でも、モンテカルロ木探索のような2手以上先を推測するアルゴリズムに大きく優位性があることを確認した。また、従来手法と教師あり学習を比較すると、教師あり学習はスコアは低いものの生存ターンや最大連鎖が大きくなっていることから、教師あり学習の特徴として生存力が高いことがわかった。

## 5 まとめ

近年のゲームAIの研究の成果により、テトリスのゲームAIの成功例が増えてきた中で、テトリス相手に特化したぶよぶよのゲームAIに有用なアルゴリズムを比較した。本研究の成果は、単なるゲームAIの開発に限定されるものではないと考えられ、様々な領域において価値を持つ可能性がある。ゲームの仕組みに合わせた能力と状況判断能力のあるようなAIは、自動運転や医療支援などの様々な分野で有益な役割を果たすことが期待される。

## 参考文献

- [1] mayah, <https://mayah.jp/article/2015/puyoai/> (参照: 2024/1)
- [2] 添島 克駿, 山口 暢彦, 福田 修, 奥村 浩. モンテカルロ木探索を用いたぶよぶよ AI. 第36回ファジィシステムシンポジウム 講演論文集 (FSS2020 オンライン), pp.527-530
- [3] 高田 宗一郎, 清 雄一, 田原 康之, 大須賀昭彦. 深層強化学習によるぶよぶよ AI. 信学技報, vol.120, no.379, AI2020-49, pp.59-64
- [4] AlphaGo, <https://deepmind.google/technologies/alphago/> (参照: 2024/1)