修士学位論文要約（令和４年３月）

領地拡大型対戦ゲームにおける深層強化学習に関する研究

山下　裕太郎
指導教員：周 暁,　学位論文指導教員：鈴木 顕

Deep Reinforcement Learning in Flood-Filling Competitive Games

Yutaro YAMASHITA
Supervisor: Xiao ZHOU,　Research Advisor: Akira SUZUKI

In recent years, a machine learning technique called deep reinforcement learning has achieved remarkable results, especially in board games. This study deals with a type of game widely studied in the field of computer science, called flood-filling competitive games. Flood-filling competitive games are games in which two players compete to expand the territory of each other by alternately performing operations of colors on a given board, and compete to see how large their territory will eventually be. In this paper, we applied deep reinforcement learning methods to one of the most basic flood-filling competitive games, the competitive Flood-It, to explore the knowledge of deep reinforcement learning in flood-filling competitive games. As a result, we concluded that it is difficult or at least time-consuming for a neural network to learn territories and territory colors by itself in learning flood-filling competitive games, and that an additional input of territories of each player is effective in improving the learning efficiency of the neural network.

1. はじめに

ゲームは勝敗という目的がはっきりしている性質から，人工知能の分野で広く研究されてきた．近年では特に，囲碁やチェス，将棋といったボードゲームで，深層強化学習と呼ばれる機械学習の手法が目覚ましい成果を上げている [3]．深層強化学習とは，ディープニューラルネットワークを用いた人工知能が，試行錯誤により自ら学習を行うような機械学習アルゴリズムの総称である．

本研究では，計算理論の分野で広く研究されている領地拡大型対戦ゲームと呼ばれる種類のゲームについて取り扱う．領地拡大型対戦ゲームとは，二人のプレイヤーが与えられた盤面に対し決められた色の操作を交互に繰り返す事で互いの領地を拡大し，最終的な領地の広さを競うゲームの総称である．

本論文では領地拡大型対戦ゲームの中で最も基本的なゲームの一つである対戦型 Flood-It[1][2]というゲームについて，深層強化学習の手法を適用し，領地拡大型対戦ゲームにおける深層強化学習の知見を探ることを目的とする．

2. 対戦型 Flood-It

対戦型 Flood-It とは，$n \times n$ のグリッド状の盤面で行う領地拡大型対戦ゲームの一種である（図 1）．

盤面における各マスには，ゲーム開始時にランダムに色が割り当てられる．ただし，一番左上のマスと一番右下のマスの色は異なることが保証される．ここで，一番左上のマスを先手のピボット，一番右下のマスを後手のピボットとして定義する．このとき，先手（後手）のピボットから同じ色のマスを上下左右にたどって到達できるマスの集合を先手（後手）の領地と定義する．図 1 において，実線で囲まれた領域が先手の領地であり，点線で囲まれた領域が後手の領地である．

対戦型Flood-Itでは，先手と後手が交互に領地に含まれるすべてのマスの色を変更する．このとき，自分と相手の領地の色は選択することができず，また領地を増やすことができる色がある場合は必ずその中から色を選ばなければならない．この操作を繰り返し，どちらかのプレイヤーが過半数のマスを領地とした時点でそのプレイヤーが勝利となる．

対戦型 Flood-It は使用する色の数が 4 色の場合であっても NP 困難であることが示されている [4]．

3. AlphaZero

AlphaZero とは深層強化学習アルゴリズムの一種であり，ディープニューラルネットワークとモンテカルロ木探索によって手を選択すること，またニューラルネットワークの学習にはセルフプレイによる対戦データのみを用いることが特徴である．

今回は，この AlphaZero を簡素化した深層強化学習アルゴリズムを対戦型 Flood-It に適用した．

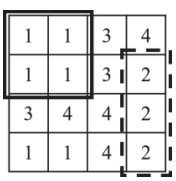

図 1　対戦型 Flood-It の盤面

4. 実験

今回の実験では盤面の大きさを 12×12，色数を 5 とした．そして，領地入力の有無，および色固定の有無という二つの条件を切り替えてそれぞれ学習させ，学習済みモデルの性能を比較した．

ここで，領地入力とは先手と後手の領地を表す追加入力であり，また色固定とは，先手の領地の色および後手の領地の色が常に同じ色となるように盤面を変換する処理である．

モデルの性能評価には，エージェントをランダムに手を選択するプレイヤー（以下ランダムプレイヤー）およびその色を選択することで広くなる領地が最大となるような色を選択するプレイヤー（以下貪欲プレイヤー）と一定回数対戦させることで行った．このとき，ひとつの初期盤面に対して先手と後手を交代してそれぞれ一回ずつ対戦を行った．これを繰り返したとき，先手と後手ともにエージェントが勝利した回数を W，ともにエージェントが敗北した回数を L として，完勝率を $W/(W+L)$と定義する．今回は $W+L=100$ となるまで対戦を行い，勝率に加えて完勝率の測定も行った．

5. 結果

実験の結果，各条件における最終的な勝率は，ランダムプレイヤー相手でおよそ 90%前後，貪欲プレイヤー相手でおよそ 60%前後となった．また，ランダムプレイヤー相手の完勝率は，どの条件でも最終的にほぼ 100%に到達した．一方で，貪欲プレイヤー相手の完勝率は条件によって大きく差がついた．したがって，各条件における貪欲プレイヤー相手の完勝率のみをまとめたグラフを図 2 に示す．

図 2 より，まず色固定の有無に関わらず，領地入力を行うことで貪欲プレイヤー相手の完勝率が明らかに上昇することが分かった．このことから，ニューラルネットワークが自力で領地を認識することは難しいか，少なくとも学習に時間がかかるということが考えられ，またニューラルネットワークの効率的な学習には領地入力が効果的であると考えられる．

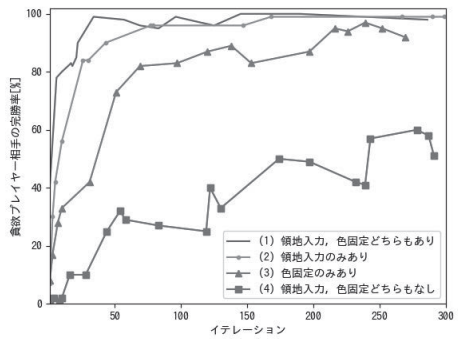

図 2　貪欲プレイヤー相手の完勝率

続いて色固定の有無に注目した場合，(3) 色固定のみありと(4) 領地入力，色固定どちらもなしの間には完勝率に大きな開きがあることから，ニューラルネットワークが自力で領地の色を認識するのは難しいか，少なくとも学習に時間がかかるということが考えられる．一方で，(3) 色固定のみありと(1) 領地入力，色固定どちらもありを比較すると，(1) 領地入力，色固定どちらもありの方が完勝率の収束がやや早いものの，最終的な完勝率にはほとんど差が見られなかった．このことから，ニューラルネットワークの学習効率を向上させるには領地入力のみを行えば十分であり，色固定はあまり重要ではないと考えられる．

6. まとめ

本研究では，AlphaZero と呼ばれる深層強化学習アルゴリズムをもとにした対戦型 Flood-It の対戦 AI を実装した．さらに，領地入力，色固定の二つの条件を変えて対戦型 Flood-It の学習を行い，実験によりその性能を比較することで，領地拡大型対戦ゲームに対する効果的な深層強化学習のアプローチを探るとともに，ニューラルネットワークが領地拡大型対戦ゲームの盤面をどのように認識しているのかを検証した．

その結果，領地拡大型対戦ゲームの学習においてニューラルネットワークが自力で領地および領地の色を学習するのは難しいか，少なくとも時間がかかるということ，また領地拡大型対戦ゲームの学習においてニューラルネットワークの学習効率を向上させるには領地入力のみを行えば十分であり，色固定はあまり重要ではないということが分かった．

文献

1) D. Arthur, R. Clifford, M. Jalsenius, A. Montanaro, and B. Sach. The complexity of flood filling games. In Proceedings of the Fifth International Conference on Fun with Algorithms, Vol. 6099 of Lecture Notes in Computer Science, pp. 307–318. Springer, 2010.
2) R. Fleischer and G. J. Woeginger. An algorithmic analysis of the honey-bee game. Theoretical Computer Science, Vol. 452, pp. 75–87, 2012.
3) D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. Science, Vol. 362, No. 6419, pp. 1140–1144, 2018.
4) 小田将也．領地拡大型ゲームにおける勝敗判定の計算複雑性に関する研究．修士学位論文，2021.