2P-01

AlphaZeroによる難易度自動調整ゲームエージェントの生成

鍋谷 優作 †　　矢吹 太朗
千葉工業大学 社会システム科学部 プロジェクトマネジメント学科 ‡

1　序論

ゲームの面白さに関する考え方の一つとしてフロー理論がある．この理論では人間がその時にしていることに，没頭している精神状態をフローと呼んでいる．フローと呼ばれる経験には少なくとも一つ以上含まれるとされる要素が八つあり，その一つに適度な難易度であることがあげられている [1]．これを踏まえて，強くなり続けるゲームエージェント（以下エージェント）を作るのではなく，プレイヤーの強さによって適切な強さに調節できるエージェントを生成する．

相手プレイヤーの強さに合わせて強さを変える研究は過去にいくつか行われている．例として将棋において相手と同程度の実力を持つ指し手に調整し，人間らしい指し手を目指すエージェントの開発や [2]，格闘ゲームにおいて相手の実力に対して動的に難易度を調整するエージェントの生成を強化学習によって行う研究などがある [3].

この研究では，DeepMind 社が開発した AlphaZero のアルゴリズムをもとに相手の実力に合わせて難易度を調整するエージェントの生成を行う．AlphaZero とは戦略に関する予備知識なしに，設定されたルールに従い物事を熟達するための汎用アルゴリズムである．

2　目的

対戦相手の強さに合わせて近い強さに調節するエージェントを自動生成する．その後生成されたエージェントを様々な強さのエージェントと対戦させその性能を調査する．

3　手法

本研究の前提として近い実力の相手と何度か対戦した場合，その対戦結果は勝率が 5 割へと収束するのではないかと考え，今回は生成したエージェントが勝率 5 割に近づくことを目標として進める．

研究は以下の手法で行う．

1. 本研究では，公開されている AlphaZero のアルゴリズムを使用したプログラム群を使用する [4].
2. 元のゲームルールはコネクトフォーと呼ばれる四目並べとなっているためルールをオセロに変更する．
3. 学習の指標となる報酬を調節し，学習サイクルを実行する．本研究では勝率が 5 割になるように調節することとする．
4. 通常の報酬で学習を行ったエージェントも用意し，学習の進行状況によってバージョン別でエージェントを複数保存する．
5. 報酬を調節したゲーム AI と通常報酬でのゲーム AI を対戦させ性能の調査を行う．

AlphaZero の学習サイクルは自己対戦，エージェント再訓練，エージェント評価によって構成されている．それぞれについては以下のとおりである．

- 自己対戦では現在最も優れているエージェント同士を対戦させエージェント再訓練に利用する学習データを作成する．
- エージェント再訓練では学習データを用いてエージェントを再訓練し新たにエージェントを作成する．
- エージェント評価では最も優れいるエージェントと再訓練したエージェントを対戦させ，勝率による評価を行い，再訓練後が勝った場合エージェントを更新する．

今回は勝率 5 割を目指すように学習するためエージェント評価では引き分けによって評価するように変更を行った．

DeepMind 社での AlphaZero では，自己対戦，エージェント評価はそれぞれ 25000 回，400 回であったが，今回使用する計算機でそれを行うには膨大な時間がかかるため，それぞれ 500 回，10 回に変更し 1 サイクルの時間を短縮している．

Generation of game agent by automatic adjustment of difficulty by AlphaZero.

† Yusaku NABETANI (s1642097nh@s.chibakoudai.jp)
‡ Department of Project Management, Faculty of Social Systems Science, Chiba Institute of Technology.

4　結果

機械学習とエージェントの性能調査の結果は以下の通りである．

- 対戦の勝敗結果のみで報酬の調節を行った結果，報酬が蓄積されず学習が滞った．ここから勝敗が決した時点での持ち石の数によっても報酬を与えるように変更を行った．
- 報酬を変更後，1 サイクル 4 時間の学習を合計 40 サイクル行いエージェントを生成した．
- 通常報酬での学習を行ったエージェントでは学習サイクルを 10 回行い最新エージェントが更新された場合に限り更新前のエージェントを保存し，9 種類のバージョンのエージェントを生成した．
- 報酬を調節したエージェントと通常報酬でのエージェントをひとつのバージョンにつき 10 回，計 90 回対戦させた結果，前者の勝率は約 57% となった．

バージョンごとの勝率，勝利時の持ち石差の平均を表 1 に，学習の進み具合を示したものを図 1 に示す．

表 1　AI 同士での対戦結果

敵バージョン	勝率	平均石差
1	0.4	12.2
2	0.6	8.9
3	0.8	11
4	0.4	12
5	0.6	9.3
6	0.5	16.9
7	0.6	14.8
8	0.6	13
9	0.6	17.9
全体	0.57	12.89

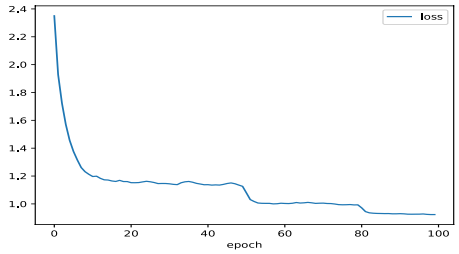

図 1　学習曲線

5　考察

表 1 によると敵バージョンと勝率に強い結びつきはなく，どの強さのエージェントと戦っても 5 割強付近であるといえる．この結果から一つの AI で敵のプレイヤーの強さによって強さを調節していると考察できる．これにより対戦ゲームにおいて AlphaZero を応用することで難易度毎にエージェントを作成するなどのエージェント作成工程の短縮を図ることができると考えた．

また図 1 によると早い段階で変化が緩やかになってしまっているため，改善の必要はあると思われる．

6　結論

AlphaZero のアルゴリズムをもとにランダムな相手に対して勝率 5 割となるオセロエージェントを自動生成し，性能を調査した．その結果勝敗での報酬を調節することで任意の強さに調節できることが分かった．

今回の研究ではエージェント相手で性能の評価を行ったが，対人での評価を行えていないため，今後は対人での評価と報酬の調整，学習の効率化をしていくべきであろう．

参考文献

[1] 小原一馬. 遊びの面白さ:遊び理論におけるゴフマン社会学の位置付け. ソシオロジ, Vol. 56, No. 2, pp. 3–118, 2011.

[2] 仲道隆史, 伊藤毅志. 人を楽しませる接待将棋システム. 人工知能学会全国大会論文集, Vol. 2014, pp. 1E5OS23b5i–1E5OS23b5i, 2014.

[3] 中川明紀, 逢坂翔太, 柴崎智哉, 柴崎智哉. ニューラルネットワークによる格闘ゲーム AI の難易度調整及び行動多様性向上手法. 全国大会講演論文集, 第 70 巻, pp. 801–802. 情報処理学会, mar 2008.

[4] David Foster. Python と Keras を使って AlphaZero AI を自作する. https://postd.cc/applied-data-sciencehow-to-build-your-own-alphazero-ai-using-python-and-keras/（2019.9.13 閲覧）.