6P-01

完全情報ゲームにおける行動価値関数を用いた不完全情報ゲームの盤面推定

共田 圭佑 †
筑波大学情報学群情報科学類 †

長谷部 浩二 ‡
筑波大学システム情報系 ‡

1 研究の背景と目的

近年、ゲームをプレイするプログラム（エージェント）を開発する試みが数多くなされている。その中で、囲碁や将棋などの完全情報ゲームにおいては、深層強化学習などの手法を用いてプロのプレイヤーを超える性能を持つエージェントも現れている。一方で、麻雀やポーカーなどの不完全情報ゲームにおいては非公開の情報が存在するため、プレイの際にゲームの状態を推測しなければならず、完全情報ゲームに対する手法をそのまま応用することができない。そのため、現在においても人間に勝利するレベルに達していないゲームが多く存在する。

本研究は、不完全情報ゲームの 1 つであるガイスター [1] をプレイするエージェントを開発することを目的とする。ガイスターはチェスのように 2 人で交互に駒を動かすゲームで、2 種類の駒の色（赤と青）が相手に非公開であるという特徴がある。なお、本研究では議論を単純化するために、ガイスターの盤面を本来の大きさである 6 行 ×6 列から 5 行 × 4 列にし、ゲーム開始時の赤駒と青駒を半分の個数である 2 個ずつにすることで、状態数の削減を行った。

ガイスターのエージェントを作成する手順として、駒の色を公開情報とした完全情報ゲームのガイスターを対象として、指し手と盤面の状態から指し手の評価値を算出する行動価値関数を深層強化学習を用いて作成する。この行動価値関数から得られる相手の指し手の評価値を用いて駒の色を推測し、推測した駒の色の尤度を元に最も価値の高い手を選択する。

本研究では作成したエージェントを、ランダムに手を選ぶエージェント、および駒の色が常に見えている状態でモンテカルロ木探索を利用し手を選ぶエージェントと 1000 回対戦させた。その結果、ランダムエージェントに対しては 730 勝 270 敗となり、モンテカルロ木探索エージェントに対しては 171 勝 829 敗となった。開発した手法により、合理的な手を選択するエージェントに対して相手の駒の色を推測できることが確認された。一方で、作成したエージェントはモンテカルロ木探索エージェントに負け越す結果となり、エージェントの性能が良いとは言えない。

2 盤面の推定と指し手の決定

本研究では、完全情報ゲームのガイスターの行動価値関数を作成する際に AlphaZero[2] で用いられたアルゴリズムを使用し、自己対戦数を 500 とし学習のエポック数を 100 とした学習サイクルを 12 セット行った。

作成した行動価値関数を用いて盤面の推定を以下のように行う。ガイスターにおいて、駒の色は非公開なものの各駒の色の総数や駒の位置は公開情報とされているため、駒の色の組み合わせを推測することで、盤面全体を推測することができる。ここで、駒の色の組み合わせを推測する際に必要な関数および要素を、形式的に次の 3 つで表す。

- C：駒の色の組み合わせの集合
- $g(x)$：実際の盤面が $x \in C$ である蓋然性 (ただし $0 \leq g(x) \leq 1, \forall x \in C, \sum_{i \in C} g(i) = 1$)
- $V_x(m)$：$x \in C$ において指し手 m を選択したときの行動価値

ここで、配置 $x \in C$ である蓋然性の有限集合 $g(x)$ を推測値と定義する。相手が指し手 m を選択したときに、行動価値 $V_x(m)$ の値を用いて $g(x)$ の値を更新する。一般に、$V_x(m)$ の値が大きいほど $g(x)$ の値が大きくなるように設定する。

Board estimation in imperfect information games using the action value function in perfect information game

† Keisuke Tomoda, University of Tsukuba, College of Information Science
‡ Koji Hasebe, University of Tsukuba, Faculty of Engineering, Information and Systems

一方で、自分の指し手を選ぶ際には、指すことが可能な各手 m について $\sum_{x \in C} V_x(m) \cdot g(x)$ を求め、この中でもっとも高い値となる手を選択する。

3 実験による評価

3.1 実験の設定

作成したエージェントを、ランダムな指し手を選ぶエージェント、およびこちらの駒の種類が常に見えている状態でモンテカルロ木探索により指し手を選ぶエージェントと 1000 回対戦させ勝敗を計測した。なお、お互いの指し手の数が 100 を超えた試合は引き分けとした。

さらに、相手の実際の駒の種類に着目して推測の精度を定義し算出した。推測の精度を定義する際に追加で必要な要素を、次の 3 つで表す。

- E：対戦相手の駒の集合
- B_p：駒 $p \in E$ が青駒である蓋然性 (ただし $0 \leq g(x) \leq 1, \forall x \in E, \sum_{i \in E} g(i) = 1$)
- $Cb(p)$：駒 $p \in E$ が青駒である、駒の色の組み合わせの集合 (ただし、$Cb(p) \subset C$)

ここで、ある配置 x について、駒 $p \in E$ が青駒である蓋然性 B_p は、$B_p = \sum_{x \in Cb(p)} g(x)$ を求めることができる。B_p は、1 に近ければ青駒であると推測し、0 に近ければ赤駒であると推測していることになり、以下の式を用いて駒 1 つに注目した推測の精度を求めることができる。

$$\begin{cases} \dfrac{B_p - 0.5}{2} & \text{(実際の } p \text{ の色が青)} \\ \dfrac{0.5 - B_p}{2} & \text{(実際の } p \text{ の色が赤)} \end{cases}$$

これを全ての駒について求め足し合わせることで、駒の色の組み合わせに対する推測が正確であれば 1 に近づき正確でなければ -1 に近づく $[-1, 1]$ の範囲の値が算出される。この値を推測の精度と定義し、対戦中の変化を計測した。

3.2 実験の結果

対戦中の推測の精度を図 1 に示す。図 1 の縦軸は推測の精度を示し、横軸はお互いの指し手の合計数を示す。図 1 を見ると、対戦相手がランダムエージェントの時よりもモンテカルロ木探索エージェントの時の方が推論の精度が高いことが確認される。この理由として、価値とは関係なく指し手を選択するランダムエージェントと比べ、モンテカルロ木探索エージェントが価値の高い合理的な指し手を選択することが考えられる。

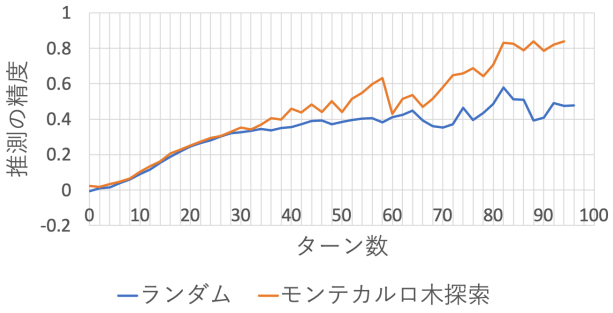

図 1 推測の精度

また、対戦の結果はランダムエージェントに対しては 730 勝 270 敗となり、モンテカルロ木探索エージェントに対しては 171 勝 829 敗となった。なお、引き分けは存在しなかった。結果を見ると、作成したエージェントがモンテカルロ木探索エージェントに対して大幅に負け越していることが確認される。この理由として、多くの対戦で序盤に決着がついてしまうことに加え、序盤の推測があまり正確でないことが考えられる。

4 結論と今後の課題

ガイスターをプレイするエージェントを構築するための駒の色の推測手法を提案した。具体的には、完全情報のガイスターについて学習した結果をもとに、行動価値関数を用いて駒の色を推測するというものである。開発した手法により、合理的な手を選択するエージェントに対して正確な推測ができることが確認された。一方で、作成したエージェントはモンテカルロ木探索エージェントに負け越す結果となり、エージェントの性能が良いとは言えない。

今後の課題として、推測値を更新する際に過去の推測値をどの程度重視するのか比率を調整し、合理的な手を選択するエージェントに対して適切な比率を求めたいと考えている。

参考文献

[1] メビウスゲームズ. ゲームリスト/ガイスター, 2011-05-17. http://www.mobius-games.co.jp/Gester.htm, (参照 2020-12-20)

[2] D. Silver et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419):1140-1144, 2018.