

# 大貧民における提出手役の学習と それを用いたモンテカルロプレイヤーに関する研究

1150293 岡 和人 【松崎研究室】

## 1 はじめに

近年、コンピュータ大貧民において、乱数によるシミュレーションを用いたモンテカルロプレイヤーが広く用いられている。モンテカルロプレイヤーでは、各盤面から乱数を用いた仮想的なゲームのプレイを行う（これをプレイアウトと呼ぶ）。より強いモンテカルロプレイヤーの実現には、プレイアウトの精度の向上が重要である。

本研究では、札譜データからの学習によって提出手役評価関数を作成し、それを用いることでプレイアウトの確度向上を試みる [1]。

## 2 提出手役評価関数と札譜データからの学習

本研究において、提出手役評価関数は相対的な手役の強さを返す関数とする。提出手役評価関数の学習アルゴリズムには 3 層ニューラルネットワーク (NN) と平均化パーセプトロン (AP) を使用し、盤面の特徴量を 182 項目設定した。

学習に用いる教師データには、既存の 4 つのプレイヤーの札譜データを使用した。それぞれ 60000 盤面の札譜データを用意し、うち最大 50000 盤面を学習に使用し、学習に使用しない 10000 盤面を評価に使用した。

評価値が最大となる手役が、実際に札譜で出された手役と一致した確率を一致率とする。各学習アルゴリズムの札譜データごとの、最大の一致率を表 1 に示す。複雑で強いプレイヤー (paoonR2) に対して、67% 程度の一致率が得られ、単純なプレイヤー (Greedy) に対しては 90% を超える一致率が得られた。モンテカルロプレイヤー (MC) に対しては 80% 程度の一致率となり、既存研究 [2] よりも高い一致率となった。また、3 層ニューラルネットワークよりも平均化パーセプトロンの方が一致率が安定している。

## 3 モンテカルロ法への適用と評価実験

単純なモンテカルロ法を用いるプレイヤーでは、そのプレイアウト中に等確率に手を選択する。本研究では、提出手役評価関数による評価値を用いてそれぞれの手に選択される確率を設定する。具体的には、この確率は評価値の高い順に等比数列となるように設定した。提案手法によるプレイヤーの強さを対戦で得られる得点によって調査した。対戦はそれぞれ 10000 ゲーム行った。単純なモンテカルロ法と提案手法のいずれもプレイアウト回数は 1500 回とした。

1 つ目に、提出手役評価関数の使用による、プレイヤーの強化を調査した。対戦の組み合わせは、提案手法によるプレイヤー 1 つと単純なモンテカルロプレイヤー 4 つと

表 1 最大の一致率

|    | paoonR2 | kishimen | MC    | Greedy |
|----|---------|----------|-------|--------|
| NN | 0.669   | 0.832    | 0.800 | 0.899  |
| AP | 0.714   | 0.903    | 0.792 | 0.947  |

表 2 モンテカルロプレイヤーとの対戦 (ベースとの差)

|    | paoonR2 | kishimen | MC    | Greedy |
|----|---------|----------|-------|--------|
| NN | +4447   | +3364    | +3822 | +3782  |
| AP | +3285   | +4051    | +3997 | +2966  |

した。得られた得点を表 2 に示す。どの場合も、単純なモンテカルロプレイヤーに対して大きく勝ち越した。一致率には差が見られたものの、プレイヤーの強化の点では明らかな差は確認出来ない。

2 つ目に、学習した札譜のプレイヤーとの対戦を行った。実験の結果、学習を行った札譜のプレイヤーとの対戦において、明らかな得点の向上は見られない。一方、例外はあるものの、強いプレイヤーから学習を行うと、得られる得点が向上した。

## 4 まとめ

札譜からの学習によって、手の相対的な強さを出力する提出手役評価関数を作成し、その性能を評価した。一致率については、平均化パーセプトロンの方がより安定して高く、より良い結果を得た。

提出手役評価関数をモンテカルロプレイヤーのプレイアウトに適用することで、プレイアウトの確度が向上し、強さが向上した。特に、対戦相手の札譜よりも、強いプレイヤーの札譜からの学習が有効であることが分かった。

モンテカルロ法を用いるプレイヤーをより強化する方法として、いくつかの知見を得た。強いプレイヤーの札譜からの学習はプレイヤーの強化に有効であったが、その提出手役評価関数の一致率は 67% から 71% であった。この一致率を向上させることで、プレイヤーの更なる強化ができる可能性がある。また、プレイアウトで手が選ばれる確率を盤面に合わせて変化させることでより強化できる可能性がある。

## 参考文献

- [1] 岡 和人, 松崎 公紀. 札譜データの学習を用いた大貧民モンテカルロプレイヤーの強化. 第 56 回プログラミング・シンポジウム (2015).
- [2] 地曳 隆将, 松崎 公紀. 大貧民における棋譜データからの提出手役評価関数の学習. 情報処理学会研究報告, Vol. 2014-GI-31, No. 15 (2014).