

## 2048 におけるニューラルネットワークプレイヤに関する研究

1180328 近藤 直季 【高度プログラミング研究室】

## 1 はじめに

「2048」は G.Cirulli が 2014 年に公開した確率的一人ゲームである。著者の先行研究 [1] では畳み込みニューラルネットワーク (CNN) に教師あり学習を用いた 2048 プレイヤを実装および評価した。その結果から教師あり学習における与える教師データ (学習) 量が不足している可能性を示唆した。

そこで本稿では、先行研究 [1] の CNN に与える教師データ量を増加させた場合の評価結果について、CNN プレイヤに「2048」をプレイさせる際の新たな方法の結果について報告する。

## 2 先行研究 [1] の CNN の構成と結果

先行研究 [1] における CNN の構成は入力層を 1 層、畳み込み層を 8 層、全結合層を 1 層、出力層を 1 層となっている。また各畳み込み層の出力は活性化関数の一つである ReLU 関数を用いる。この CNN の入力には「2048」の現盤面、出力は最善手 (上下左右のうち 1 方向) である。

平均得点 30 万点のコンピュータプレイヤのプレイログを教師データとして 1 億局面 (1 局面は 1 盤面と盤面に対する入力方向) を与えて学習させた。1 億局面の学習が完了した CNN に「2048」を 1 万試合プレイさせて評価した結果、平均得点 16,634 点、クリア率 27.82% となった。また、500 万局面の学習が完了する毎に評価した結果、平均得点は右肩上がりに増加することが確認でき、ここから 1 億局面では教師データ量 (学習局面数) が不足していると考えた。

## 3 実験

先行研究 [1] と同様の平均得点 30 万点のコンピュータプレイヤのプレイログのうち、重複なしで無作為に 10 億局面分 (先行研究 [1] の 10 倍) 用意した。1 億局面の学習に約 50 時間掛かるため用意した分の学習が完了していないが、1 億局面毎の学習が完了した際の CNN で「2048」を 1 万試合プレイさせた結果は表 1 の通りである。クリア率はゲームオーバー時に少なくとも 1 回は 2048 (クリアライン) 以上のタイルに到達した割合を指す。学習局面数を増加させた結果、1 億から 7 億では平均得点は 1.6 倍増加し、クリア率は 20% 増加した。さらに学習局面数を増やした場合に平均得点とクリア率が向上するかは不明であるが、向上する可能性は高い。

次に「2048」では、ある盤面を  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  だけ回転させた盤面、またそれらの鏡像である盤面は、元の盤面と本質的に等価である。つまり、1 つのある盤面から回転と鏡像によって 8 つの等価な盤面が作成できる。

表 1 学習局面数による性能の変化

学習局面数	平均得点	クリア率
1 億局面	16,635.9	27.63%
2 億局面	19,913.2	35.17%
3 億局面	22,445.9	41.37%
4 億局面	22,364.5	39.76%
5 億局面	22,324.2	40.33%
6 億局面	26,143.9	47.55%
7 億局面	26,680.7	48.28%

表 2 等価な 8 盤面を用いたプレイ方法の結果

学習局面数	平均得点	クリア率
1 億局面	33,760.9	62.78%
2 億局面	38,386.0	66.06%
3 億局面	42,232.5	69.29%
4 億局面	44,231.0	69.13%
5 億局面	44,832.7	70.51%
6 億局面	46,085.7	71.14%
7 億局面	49,577.5	74.04%

この特性に注目し、学習させた CNN で「2048」をプレイさせる際に盤面 (1 つ) を入力し、1 つの出力で盤面に対する入力方向を決定する方法から、盤面を 8 つ作り、入力を 8 回行い 8 つの出力の多数決で入力方向を決定する (表 1 と同じ CNN プレイヤで異なるプレイ) 方法に変更した。この結果は表 2 の通りである。学習局面数 7 億では従来 (表 1) のプレイ方法と比べ、平均得点は 1.85 倍増加し、クリア率は 25% 増加した。また、これまで 8192 のタイルに到達することが限界であったが、一段階上の 16384 のタイルに到達できるようになり、最高得点は 291,716 点に達した。

## 4 まとめ

先行研究 [1] の CNN に対して教師データ量 (学習局面数) を増加させた結果、性能向上が確認できた。また、学習した CNN で「2048」プレイ時に等価な 8 盤面を用いる方法により、さらなる性能向上が確認できた。今後は、さらに学習局面数を増加させ、性能向上の上限を検証する必要がある。仮に上限があれば、先行研究 [1] の CNN の構成を根本的に改善する必要がある。

## 参考文献

- [1] 近藤直季, 松崎公紀, “2048 におけるニューラルネットワークプレイヤの育成報告”, 第 59 回プログラミングシンポジウム予稿集, pp.95-104, 2018.