

ゲーム 2048 における対称性サンプリングの利用による N タプルネットワークの性能向上の検証

1230380 樺山 飛彗 【高度プログラミング研究室】

1 はじめに

「2048」は2014年にG.Giruliによって公開された確率的一人用ゲームである。2048は「ルールは簡単だが、極めるのが困難」といった特徴を持っており、コンピュータプレイヤーとしてNタプルネットワークを用いるもの、ニューラルネットワークを用いるものなどの研究が行われている。現在の主流であるNタプルネットワークを用いたプレイヤーの研究では、対称サンプリングを用いることで重みの数を削減でき、それにより性能が向上すると考えられている。だが対称サンプリングが学習性能の上昇につながっているとしたSzubert[1]の論文では対称サンプリングによる性能向上は正確に検証されていない。本研究ではタプルネットワークプレイヤーにおける、盤面の対称性の利用有無によって性能が向上するのかを検証していく。

2 Nタプルネットワークプレイヤーと 対称性サンプリング

Nタプルネットワークプレイヤーは盤面のN個のマス（以降Nタプルと呼称）を複数組み合わせ、それらの部分評価関数の線形和を盤面の評価値とする手法である。この手法では評価値の計算が $16^n \times$ タプル数となり、タプルの長さとNタプルの数の増加により学習の際の重みの数が莫大になる。これに対する対策として「2048」の盤面が対称的であることを利用してタプルの組の数を減らし、プレイヤーの性能を落とさず重みの数の削減を図ったのが対称サンプリングである。

3 実験方法

今回の実験では2種類の6タプルネットワークプレイヤーを作成し、合計のサンプリング回数が一致するようにそれぞれのNタプルの数を調整し、計算時間ごとのスコアを比較する。以下が比較するプレイヤーである。

- 6タプル×4組の対称サンプリングを用いたプレイヤー
- 6タプル×32組の対称サンプリングを用いないプレイヤー

それぞれサンプリング数は32回となっている。部分評価値の補正に関してはTD-AFTERSTATEによる学習を用いる。どちらも2000万回分ゲームの学習を行い、経過時間と実行回数、平均スコアを比較する。

4 実験結果

学習から約10時間半経過した結果を表1に示す。

表1 38200秒経過時点での試合回数と平均スコア

	試合数	平均スコア
対称性利用あり	20000000	214300
対称性利用なし	18880000	216560

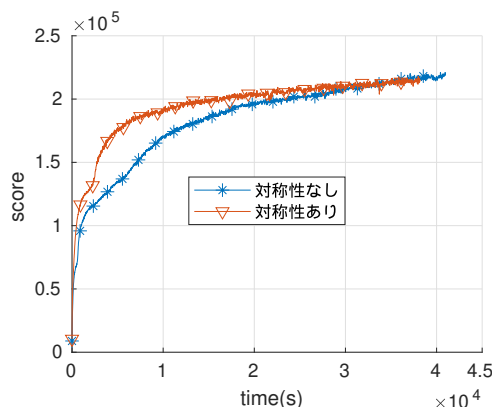


図1 経過時間と平均スコア

結果としては試合数は対称性利用のプレイヤーが110万試合ほど多く、同程度の時間が経過した際の平均スコアは対称性不使用のNタプルネットワークプレイヤーが対称性を使用したプレイヤーをわずかに上回る結果となった。グラフを見ると初期のスコアは対称性ありの方が高いが、最終的にどちらも近いスコアになっている。

5 まとめ

本稿ではNタプルネットワークプレイヤーにおける対称サンプリングを使用することにおける性能向上を検証した。サンプリング回数を同条件にした場合、対称サンプリングを利用した重みの数の削減による計算時間の短縮は大幅な効果は見られず、Nタプルネットワークプレイヤーの時間当たりの性能向上につながっているとは言えないことが分かった。

論文の本体で、対称性のありなしで盤面の評価値に変化があるかについても議論する。

参考文献

- [1] Marcin Szubert, Wojciech Jaskowski, Temporal Difference Learning of NTuple Networks for the Game 2048, IEEE Conference on Computational Intelligence and Games, 2014