

ミニ2048におけるNタプルネットワークの対称性サンプリングの影響分析

樺山 飛隼 【高度プログラミング研究室】

An Analysis of Symmetry Sampling in N-Tuple Networks for Mini2048

MOMIYAMA, Hisui 【High-Level Programing Lab.】

1 はじめに

「2048」は2014年にG.Giruliによって公開された確率の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。Nタプルネットワークを用いたプレイヤーの研究では、対称性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。だが対称性サンプリングを最初に採用したSzubert and Jaskowskiの論文[1]では対称性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。著者の先行研究[2]では学習量が非常に多いところを除いて、対称性サンプリングによる性能向上が確認したが、対称性利用ありの場合でも平均スコアが 2.14×10^5 と最先端のプレイヤーの平均スコア[3]の約3分の1程度となっており、十分な分析ができていない。本研究では完全解析が済んでいることを利用し、ミニ2048環境を用いて、Nタプルネットワークプレイヤーにおける対称性サンプリングの有無が学習効率および最終性能に与える影響を定量的に評価する。

2 ミニ2048

本研究で扱うミニ2048は 3×3 のボード上でプレイされる2048の小型版である。ボードのサイズ以外にルールは変わらない。ミニ2048はこれまでに完全解析されており、すべての局面からの最善プレイにおける期待得点が求められている。

3 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個のマス組（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。学習は自己対戦によって得られる遷移 (s_t, r_t, s_{t+1}) に基づき、各種学習法を用いて、評価値を更新する。

対称性サンプリングでは、盤面の回転・反転により得られる複数の等価盤面を用い、(i)更新対象のタプル入力を対称変換した複数状態へ拡張する、または(ii)対称変換を通じて同一重みを参照するように設計することで、

見かけ上の学習データ増加と重み数削減を両立する。

4 実験条件

本研究では対称性の利用有無以外の条件を統一し、効果を分離して評価する。本研究では評価時は8対称のタプル評価値を合算し、更新時は8対称の各パターンすべてに同じ更新を適用した。

本研究では2つのNタプル（NT6とNT4b）に対し、それぞれ対称性利用の有無（symとnotsym）によりプレイヤー4種を作成した。以下が比較するNタプルの定義である。

| Name | Tuples | Number of weights |
|------|--------|-------------------|
| NT6 | | 7,086,244 |
| NT4b | | 87,846 |

図1: Nタプルの配置

symとnotsymの間で一度当たりの盤面評価値更新回数が同一になるよう調整している。NT4bの場合は24回、NT6の場合は16回重みの調整に関しては学習率を自動的に調整するTemporal Coherence学習法を用い、以下の手法によって強化する。

- マルチステージング：ゲームの前半と後半の2ステージに分けて評価値をテーブルとして格納する。
- 楽観的初期化：すべての評価値の初期値が1200となるよう初期化する。

これらの方法を用いて 5×10^8 回の重み更新を行った。

5 実験結果

5×10^8 ステップ訓練した後に100ゲーム行った際の平均と標準偏差を表1に示す。

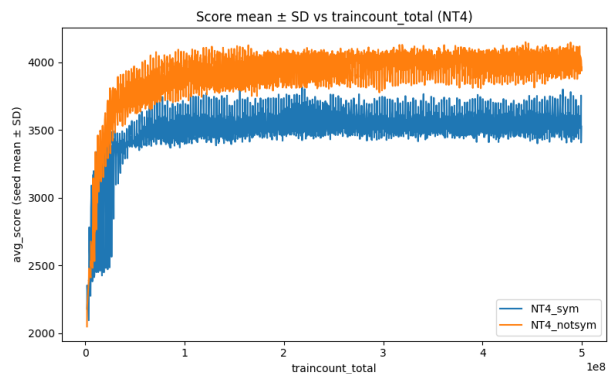


図2: トレーニングの進捗と平均スコア NT4(sym/notesym)

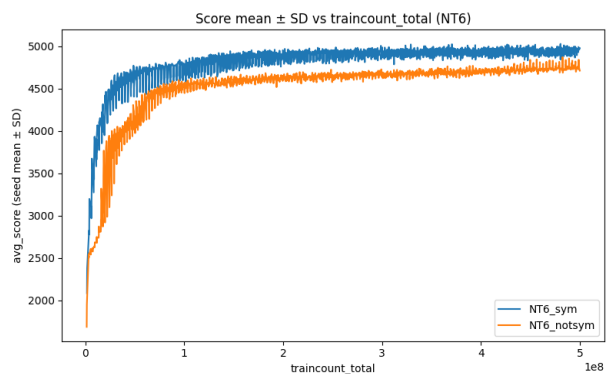


図3: トレーニングの進捗と平均スコア NT6(sym/notesym)

図2と図3はそれぞれ、学習中のエピソードのスコアを移動平均10000で求めたものである。結果として6タプルからなるNT6では対称性サンプリングありの方がより良いスコアを出しており、図3のグラフを見る限り学習も早かった。

逆に4タプルからなるNT4bを用いたものでは対称性サンプリングなしの方が、より良いスコアを出している(図2)。これらはどちらも学習を通して一貫して優位が分かれている。

6 まとめ

本稿ではミニ2048環境を用いた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影響を分析した。ミニ2048では比較的小さる4タプルを用いる際には対称性利用を使用しない方が強く、比較的大きな6タプルを用いる際には対称性を利用する方が高い性能を得た。

論文の本体では、対称性の有無で盤面の評価値分布に変化があるか、完全解析で得られている評価値を用いた散布図、ゲームの進行度(progress)に対する生存率、を用いた比較分析を行う。

参考文献

- [1] Szubert, M., Jaśkowski, W.: Temporal Difference Learning of N-Tuple Networks for the Game 2048.

IEEE Conference on Computational Intelligence and Games (2014).

- [2] 樺山飛彗: 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. 学士論文、高知工科大学(2024).
- [3] Guei, H., Chen, L.P., Wu, I.-C.: Optimistic Temporal Difference Learning for 2048. IEEE Transactions on Games 14(3), 478–487 (2022).