2048 の N タプルネットワークプレイヤに対する教師あり学習に関する研究

1180371　藤田　竜貴　【 高度プログラミング研究室 】

1　はじめに

ゲーム情報学において，ゲームを上手にプレイするコンピュータプレイヤを作成することは目的の 1 つである．ゲーム「2048」における N タプルネットワークプレイヤの強さを左右する要素の 1 つに，学習の手法がある．これまで，様々な強化学習の手法が提案されてきたが，教師あり学習の研究は比較的されておらず．教師あり学習がどれほど有効であるかは明らかではない．

そこで本稿では，松崎が提案した後退 TC 学習 [1] を用いた N タプルネットワークプレイヤと，それと同じタプルを用いたプレイヤに教師あり学習を適用したプレイヤで性能比較を行い，実験結果から考察を行う．

2　既存手法

2.1　N タプルネットワークプレイヤ

盤面のある n マスの組を N タプルという．ある盤面とその盤面の回転・鏡像変換によって得られる盤面について，タプルに該当するマスに置かれたタイルの組み合わせにより特徴量を算出し，特徴量の和によって盤面評価値を計算し，行動選択するプレイヤを N タプルネットワークプレイヤという．特徴量は，タプルに該当するマスに置かれたタイルの組み合わせに対応づけられ，特徴量テーブルに格納される．

N タプルネットワークの拡張に，多段階化（Multi-staging）がある．盤面に存在する最大のタイルによって参照する特徴量テーブルを変える手法である．本稿では最大のタイルが 256 以下，512，1024，2048，4096，8192，16384，32768 以上の 8 段階で多段階化する方式を採用する．

2.2　後退 TC 学習

時刻 t における盤面を S_t，S_t の盤面評価値を $V(S_t)$，S_t からランダムに出現したタイルを除いた時刻を $t'-1$，S_t から行動選択した直後の時刻を t'，プレイヤの行動選択によって得られた得点を r_t とする

$V(S_{t'-1})$ と $V(S_{t'})$ の差に r_t を足したものを誤差 Δ とし，その誤差 Δ が小さくなるように特徴量のテーブルを更新する．学習率 α にはこれまでの誤差の累積と誤差の絶対値の累積を用いる．

この更新手法を用いてゲームが終了した時点からゲーム開始時点まで遡りながら学習を行う方法を後退 TC 学習という．

3　提案手法

3.1　提案手法 A

後退 TC 学習では，プレイヤが行動を選択し，ゲームが終了した時点で，その結果をもとに学習を行うのに対し，この手法では，教師データと全く同じプレイをさせ，後退 TC 学習と同じ更新方法を用いて，特徴量のテーブルを更新し，学習を行う手法である．

3.2　提案手法 B

手法 A と後退 TC 学習を，ある盤面数 m 回のテーブル更新毎に切り替えて学習を行う手法である．ゲームの途中で m 回のテーブル更新が完了したとき，学習手法は切り替わらず，ゲームが終了した時切り替わる．実験では $m = 10^5$ を使用する．

4　実験結果

表 1　15×10^8 盤面学習後の各プレイヤの得点

	多段階化なし		8 段階化	
	greedy	expect	greedy	expect
後退 TC 学習	164599	268744	163189	282872
提案手法 A	4589	52618	4725	54091
提案手法 B	128859	320016	130378	348536

教師データには約平均 45 万点のデータを使用した．各学習手法に 15×10^8 盤面を学習させた後，greedy 法で 1000 ゲーム，expectimax(3–ply) 法で 100 ゲームプレイさせ，得られた平均点を表 1 に示す．

5　まとめ

本研究では，既存手法のアルゴリズムに教師あり学習適用する手法 A，B を提案し，既存手法を含むそれぞれのプレイヤで性能比較を行った．既存手法に対し，手法 A は大きく下回り，手法 B は greedy 法では下回り，expectimax 法では上回った．

手法 A ではゲーム序盤の学習がゲーム後半の学習に比べ少なく，序盤でゲームが終了してしまうことが考えられる．手法 B ではゲーム序盤の学習を強化学習で補い，ゲーム後半の学習を教師あり学習で効率的に行うことができたので expectimax 法では既存の手法を上回ることができたと考える．

参考文献

[1] Kiminori Matsuzaki, “Developing 2048 Player with Backward Temporal Coherence Learning and Restart”, 15th International Conference on Advances in Computer Games. 2017.

[2] Fujita Ryuki, Kiminori Matsuzaki, “Improving 2048 Player with Supervised Learning”, 6th Intenational Symposium on Frontier Technology. 2017.