ニューラルネットワークと強化学習による対戦型 2048 プレイヤの作成

1210389　横山智洋　【 高度プログラミング研究室 】

1　はじめに

「対戦型 2048」は，2014 年に G.Cirulli が公開した一人ゲーム「2048」を二人ゲームに拡張したゲームである．これまでの「2048」および「対戦型 2048」の研究の主流は N タプルネットワークを用いたものであった．また，近年では「2048」プレイヤの作成にニューラルネットワークを用いる研究も成果を上げている．そこで本研究では,「対戦型 2048」プレイヤの学習にニューラルネットワークによる強化学習を行い，攻撃側プレイヤと防御側プレイヤを交互に学習する方法と，同時に学習する方法の 2 種類で作成した．その後，作成したプレイヤ間で相互対戦を行うことにより評価を行った．

2　ニューラルネットワークの構成

本研究では防御側プレイヤと攻撃側プレイヤともにニューラルネットワークによる評価関数を用いて作成した．入力データには「2048」の盤面データを用いた．これらのネットワークは全 5 層からなり畳み込み層 2 層，全結合層 3 層で構成され，第 1 層から第 4 層の出力は活性化関数の ReLU 関数を用いている．第 5 層はポリシーネットワークでは出力層において Softmax 関数を用い，防御側プレイヤの手の選択確率を表すようにした．またバリューネットワークでは与えられた局面における評価値を出力するようにした．

3　実験方法

本実験では,「対戦型 2048」プレイヤの学習方法としてシナリオ A，B，C の 3 つを用意し，各シナリオについて 168 時間の学習を行った．

シナリオ A では初期プレイヤとして教師あり学習を行ったポリシーネットワークによる防御側プレイヤを用い，そのプレイヤ相手に攻撃側を 24 時間学習，次に防御側を 24 時間学習というように交互に学習を行った．このプレイヤは，通常の「2048」において平均得点 71014 点を達成するプレイヤである．教師データは Matsuzaki [2] による 2048 プレイヤのプレイログを用い，正解データはそのプレイヤの手の選択とした．

シナリオ B では，初期プレイヤとして通常の 2048 を対象としたバリューネットワークの強化学習による防御側プレイヤを用いた．このプレイヤは，通常の「2048」において，平均得点 91590 を達成するプレイヤである．その後の攻撃側，防御側の学習についてはシナリオ A と同様の方法で行った．

シナリオ C では，ニューラルネットワークの重みの初期値をランダムとして，攻撃側と防御側を同時に学習した．

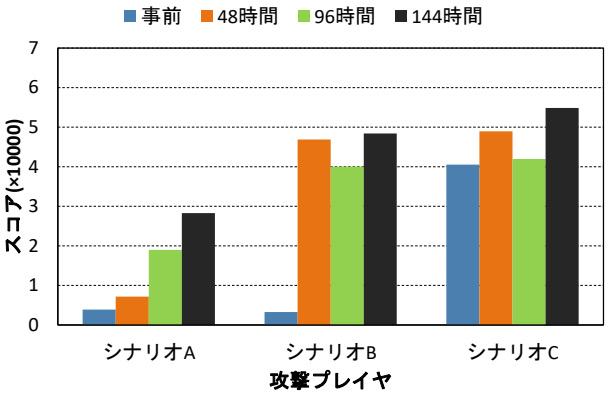

図 1　シナリオ A の防御側プレイヤとの対戦結果

4　実験結果

図 1 に作成したプレイヤ間の対戦結果を示す．対戦スコアは，同じシナリオ間の対戦ではスコアが低く，異なるシナリオ間での対戦ではスコアが高くなっていた．また，シナリオ C で学習した攻撃側プレイヤはシナリオ A，B で学習したプレイヤよりもスコアを下げられていなかった．また，初期プレイヤとシナリオ A，B で学習した攻撃側プレイヤとの対戦はどちらもスコアが 5000 点以下と，低く抑えられていた．

5　おわりに

本研究では,「対戦型 2048」のプレイヤの作成に，ニューラルネットワークによる強化学習を行い，攻撃側と防御側プレイヤを作成した．その後，作成したプレイヤ間での相互対戦によるプレイヤの評価を行った．その結果，同時にプレイヤを学習するよりも交互に学習する方がより良いプレイヤを作成できることが分かった．また，学習が特定のプレイヤに偏っていたため複数プレイヤを同時に扱ったり，プレイにランダム性を導入するなどして改善できると考えている．

参考文献

[1] 横山智洋，松崎公紀，“ニューラルネットワークと強化学習による対戦型 2048 プレイヤの作成”，第 62 回プログラミングシンポジウム予稿集，2021.

[2] Matsuzaki, K.: Developing 2048 Player with Backward Temporal Coherence Learning and Restart, *Proceedings of Fifteenth International Conference on Advances in Computer Games (ACG2017)*, pp. 176–187 (2017).