論文要旨

ゲーム「2048」のプレイヤについて

学際科学科　総合情報学コース
08-152021
金澤望生
指導教員　山口和紀教授

本研究では，強化学習を用いて訓練したパズルゲーム「2048」のプレイヤに対し，人間が 2048 をプレイする際に用いる経験的な方策を強化学習とは独立に導入することで，より優秀な成績を収められるように改良することを目指す．

2014 年 3 月に公開されて以来，そのシンプルで直感的に理解しやすいルールと攻略の難しさから，2048 は世界中の人々にプレイされるゲームとなった．一方で，2048 は完全情報ゲームでありながらランダム要素のある非決定論的ゲームであるという特徴を持ち，ゲーム AI の分野でもしばしば題材として取り扱われてきた．

Szubert & Jaśkowski (2014)は初めて強化学習を用いて 2048 プレイヤを実装した．彼らは強化学習の一種である TD 学習（TD(0)）と n タプルネットワークによる状態の近似を組み合わせてプレイヤを実装し，後の研究の多くもこれに倣ってプレイヤの実装を行った．いくつかの研究者グループにより，TD の学習プロセスのステップ分け，n タプルネットワークの配置の改良，補助的な Expectimax 木探索の導入などの改良が重ねられてきた．最近の研究では，Jaśkowski (2017)が学習率を自動的に決定する Temporal Coherence の導入や，さらなる Multi-Stage TD 学習の改良，学習プロセスにおいてよりゲーム終盤の状態を学習できるようにする Carousel Shaping の導入などを行い，平均スコア 609,104 を達成するなど非常に良い性能を記録している．

一方で，2048 には人間がプレイする際に経験的に得られた知見も存在している．例えば「盤面上で大きな数の書かれたタイルは端の 1 辺に集める」といった方策があり，先行研究の強化学習を用いたプレイヤにおいてもこの方策は習得されていることが確認された．一方で，「最も大きな数の書かれたタイルは盤面の隅に固定する」という方策は人間がプレイする際によく用いられるものの，強化学習によるプレイヤでは必ずしも習得されていると言えないことがわかった．

そのため，本研究ではこの方策を「maximum-in-corner 方策」と名付け，この方策に従っている状態の評価値に対して「corner ratio」という定数を掛け合わせ，方策に従っている状態の評価値を割り増して評価することにした．この方策とさまざまな条件を組み合わせて実験を行った結果，学習率を 0.0025 とし，学習中および検証のゲームプレイ中に方策を採用し，ゲーム中のスコアが 80,000 を超えたら corner ratio を変更するという条件において，本方策を使用しない場合よりも良い成績を得た．10 万ゲーム学習した時点で本方策を使用しない場合の平均スコアは 111,486 だったのに対し，本方策を使用した場合の平均スコアは 148,502 であった．

本研究によって，maximum-in-corner 方策の適用により TD 学習のみを用いた 2048 プレイヤの性能がさらに改善することがわかった．しかしながら，corner ratio の決定にあたっては「ゲームが進んだら corner ratio の値を下げると良い」という抽象的な傾向しか判明しておらず，本研究で行ったものよりもより厳密な corner ratio の決定方法の存在も考えられる．また，maximum-in-corner 方策以外にも，強化学習ベースの 2048 プレイヤが学習を通じて習得できず，なおかつプレイヤに組み込むことが有効な方策が存在する可能性もある．今後の研究においては，本研究において導入した maximum-in-corner 方策のより精緻な形での導入や，本方策以外の新たな経験的方策が 2048 プレイヤに対して導入されることを期待する．