卒業論文

ゲーム「2048」のプレイヤについて

08-152021　金澤望生

指導教員　山口和紀 教授

2018 年 1 月

東京大学教養学部学際科学科総合情報学コース

概要

インターネットブラウザやスマートフォン上で遊ぶことのできるパズルゲーム「2048」をプレイする AI の改良を行った．改良には盤面上で最も大きな数のタイルが隅にあることを重視する独自のヒューリスティック「corner bonus」を使用した．(仮)

キーワード　ゲーム AI，機械学習

目次

第 1 章

導入

モチベーションや 2048 の基本ルール・指標について説明します.

第 2 章

先行研究の紹介

既存研究が使用している手法とプレイヤの成績について説明します.

2.1 Szubert & Jaskowski (2014)

Szubert & Jaskowski は，TD 学習を用いたプレイヤの訓練と n タプルネットワークを用いた価値関数の表現を組み合わせることによって，人間の知識やゲーム木探索を使用しないで十分強い 2048 プレイヤを実装することに成功した。

2.1.1 TD 学習

TD 学習の「TD」とは temporal difference の略であり，すなわち状態間における価値の差分を学習することによって学習器の訓練を行う手法である。2048 にあてはめると，とある盤面 s' の価値と，その盤面の 1 プレイ後の盤面 s'_{next} の価値の差分を取り，これを現状定まっている s' に足し込んでいくことで訓練を行うことになる。TD 学習にはさまざまな派生があるが，Szubert & Jaskowski が使用している TD(0) 学習は以下の式によって表現される：

$$V(s) \leftarrow V(s) + \alpha(r + V(s'') - V(s))$$

この式において，V は価値関数，α は学習率，r は報酬である。学習率は計算された差分を価値関数の更新にどれほど反映するかを決定するパラメータである。

TD 学習は Tesauro によるバックギャモンへの適用でよく知られるようになり，碁やオセロ，チェスにおけるゲーム AI の方策決定の手法として用いられるようになった。

2.1.2 n タプルネットワーク

TD 学習によって盤面の評価とその学習を行うことができるが，盤面と評価値をどのように結びつけるかが問題になる。まず，2048 で有り得るすべての盤面に対して評価値を与える 1 対 1 対応のルックアップテーブル（LUT）を作成することを考えると，2048 で有り得る盤面の数は $(4 \times 4)^1 8 \approx 4.7 \times 10^2 1$ と膨大な数になり，このような LUT を計算機上で実装するこ

とは現実的に不可能である。

そこで，一部のマスの組み合わせによる「タプル」というクラスターを作成し，さらに複数のタプルを組み合わせることで盤面を表現する手法「n タプルネットワーク」を 2048 に導入することが，Szubert & Jaskowski によって提案された。たとえば，下記のような n タプルネットワークを実装した場合，1 つのゲーム内で保持すべき重みの数は 860625 であり，全ての有り得る盤面に対する LUT を保持するのに対して非常に少なくて済む。

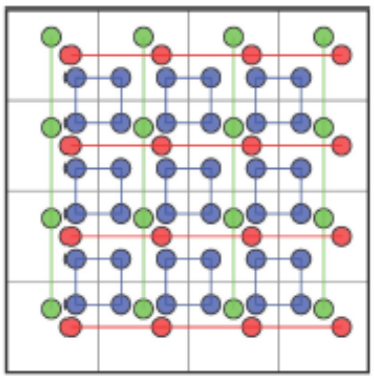

図 2.1. n タプルネットワークの例

n タプルネットワークは Bledsoe & Browning (1959) によりパターン認識に用いられたのが最初の採用例である。ゲーム AI の分野では Jaskowski (2014) によってオセロに適用され，一定の成果が得られた。

第 3 章

本研究のアイディア

本研究で導入しようとしている手法のアイディアについて説明します.

第 4 章

提案と実装

前章で説明したアイディアの具体的な提案とその実装方法を説明します.

第 5 章

実験

提案したアイディアの実験結果と既存研究の実験結果を比較します.

第 6 章

考察と結論

実験結果をもとに，結果の考察を行い，本研究をまとめます.

謝辞

謝辞を書きます.

参考文献

[1]
[2]

付録 A

表やプログラムリストの掲載が必要になったらここに掲載します.