

【要旨】

オセロの棋譜データによる教師あり学習とAI同士で対戦させる強化学習、 $\alpha\beta$ 法と組み合わせた先読みを行うことで、より強いAIのプログラム「E-Minor」を作成した。その結果、オセロの定石を再現することができ、いくつかのオセロプログラムに勝利し、人との対戦で勝率が85%を超えるなど、人間よりも強いAIを作成することができた。

【研究背景】

2015年10月、人間のプロ囲碁棋士にAIが初めて勝利した。このとき使われたのは、AlphaGoというプログラムだ。この技術を応用することで、最強のオセロAIを作ることができないだろうか。

【研究方法】

ニューラルネットワークを用いて、以下の手法でE-MinorというオセロAIを作った。

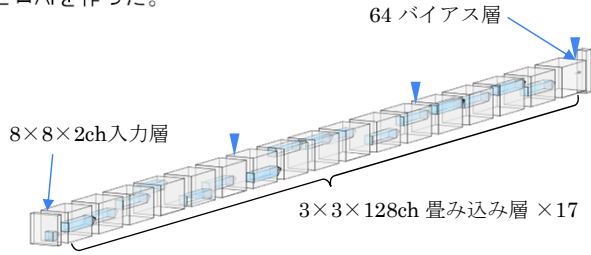


図1 E-Minorのニューラルネットワークの構造（青三角はドロップアウト層の場所）

1. オセロの棋譜（WTHOR）をダウンロードし教師あり学習を行った。
2. 1で予測率が最も高かったもの同士を対戦させ、強化学習を行った。
3. 定石が再現できるかどうかを調べた。また、既存のオセロプログラム5種や人との対戦データを集めた。
4. 最善手を探索する $\alpha\beta$ 法を用いた先読みと組み合わせた。

【結果・考察】

1. 教師あり学習

層数やフィルター数、層の種類を変えた。正解率との関係は以下の通りだ。

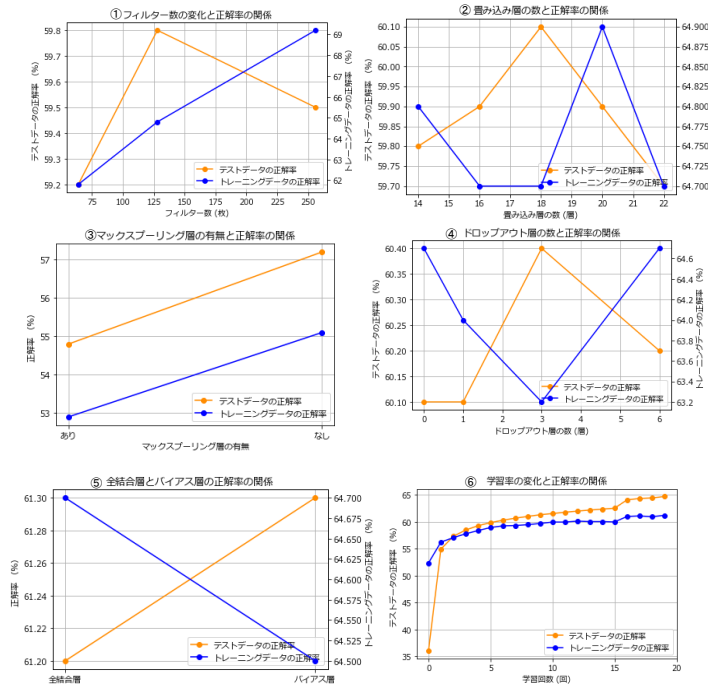


図2 条件を変えた学習による正解率の変化

図2からフィルター128枚、畳み込み層18層、ドロップアウト層3層、全結合層の代わりにバイアス層を入れる、学習率を学習回数15回以下のとき0.001、18回以下のとき0.0001、20回以下のとき0.00001にすると、正解率が高いAIを作ることができると分かった。

2. 強化学習

教師あり学習で最も強かったAI同士を対戦させ、黒を学習させたときの黒の勝率は以下ようになった。

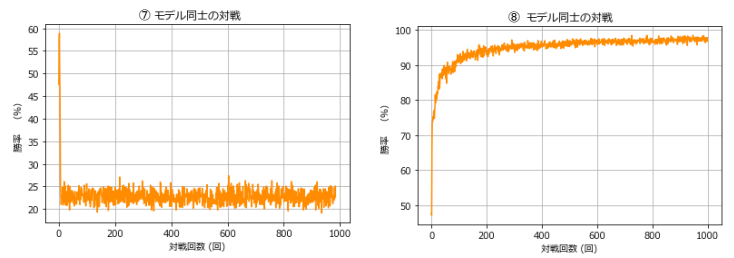


図3 学習による正解率の変化（1回のプレイで10回（左）1回（右）学習させた）

図3から、1回のプレイにつき1回学習した方が良いということが分かる。

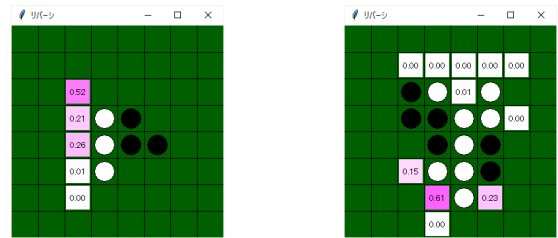


図4 虎定石を再現した盤面（数値は行動価値を表す）

E-Minorと定石を比較すると虎定石と完全に一致した。伝統的な定石と一致したことから、E-Minorは強い状態価値の計算ができていている可能性が高い。

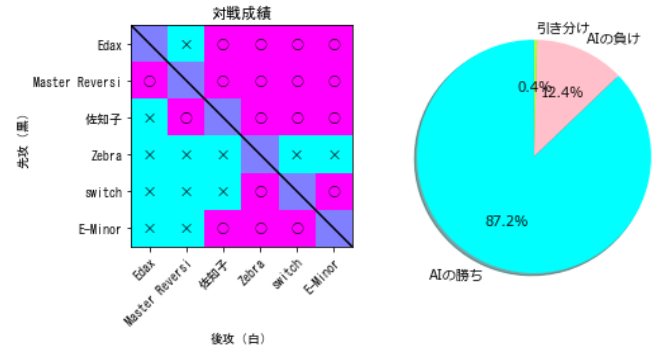


図5 プログラムとの対戦成績

図6 人との対戦成績

Master Reversiのような強豪ソフトには負けたものの、人との対戦では約87%勝利したという結果が得られた。



図7 $\alpha\beta$ 法やDeep Learningにかかる時間

43手目以降の場合、 $\alpha\beta$ 法の計算が現実的にできることが分かった。また、 $\alpha\beta$ 法と組み合わせたものとDeep Learningを対戦させたところ、 $\alpha\beta$ 法と組み合わせたものの勝率は100%だった。しかし、Master Reversiには負けてしまった。

【結論・今後の展望】

一般的な定石を再現でき、人間との対戦も約87%勝利できたことから、E-Minorが人間よりも強いオセロプログラムだと考えられる。今後は、 $\alpha\beta$ 法と組み合わせたもの同士を強化学習させ、より強いプログラムにしたい。

【引用文献・参考文献】

David Silver, Aja Huang, et al. Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587):484-489, 2016.
 布留川英一 (2019). 『AlphaZero 深層学習・強化学習・探索』. ボーンデジタル.