

夏風のアピール文章

LeelaZero の棋譜を使ってニューラルネットワーク(NN)を再学習させています。

- ・ Policy, Value の他に投了する手数、も予測する。
 - ・ 投了する手数は目数差に変換して勝率+地を多くとろうとする探索にする
 - ・ ニューラルネットワーク(NN)に入力するデータは石の配置、ダメの数、シチョウ探索の結果、その位置から 8 方向に空間が連続するか、黒石が連続するか、白石が連続するか、を与える。
- これは 3x3 のフィルタだと遠い場所の把握が苦手なようなので、それを軽減する狙いです。
- ・ playout は Aya を用いて MCTS で Value の値を組み合わせる。

以下は開発途中のメモです。

昨年の AI 竜星戦では LeelaZero(LZ)の棋譜を使って、シチョウや地合の情報を再学習させる、ということを行いました。

ただ一番の失敗は LZ が次の 1 手でなく、着手の分布を使って学習していることを理解しておらず、また

<http://zero.sjeng.org/>

で公開されている「Raw SGF files」は SGF ファイルのみで、しかも当時 LZ は序盤の 30 手まで着手回数をもとにランダムに選ぶ、でなく 999 手目まで、つまりすべての着手をランダムに選ぶ、で棋譜を生成していました。

これは 2017 年に出た AlphaZero 論文では AlphaGoZero と違い、すべての手をランダムで選んでいる、と読めるために変更したと思われます。同時期に学習を行った ELF も最初の 30 手でなく同様にすべての手をランダムで生成しています。

このため ELF は本来到達するべき AlphaGoZero の棋力よりは低い棋力になってると思われれます。

すべての手をランダムに選ぶと棋力は GNU Go 程度にまで下がり、対局結果(Value)の精度が下がっていたと思われれます。Policy は着手分布からの学習なのでそこまで影響はなかったかもしれません。

この LZ の間違いは 2018 年のサイエンス誌に掲載された AlphaZero 論文に付属の pseudocode.py が公開された時点で修正されました。

前置きが長くなりましたが、つまり昨年学習させた LZ の棋譜の Policy は非常に弱いもので出来たネットワークも当然弱いものでした。

LZ は学習で実際に使う分布のデータも公開しており、今回はこちらを使用しました。

<http://leela.online-go.com/training/>

Raw training data

<https://github.com/leela-zero/leela-zero/issues/167>

おそらく

train_6ee288bd.zip なら LZ の 249 番のネットワーク

249 2019-10-22 10:21 6ee288bd

で作った棋譜が入っていると思われます。

LZ 249 番目から 200 番目の合計 482 万棋譜を使用しました。

厄介だったのが元データは実際に学習に使う過去 7 手前までの局面を含む生データの形式で、圧縮された状態で 250GB、全部展開すると 3TB ぐらいになり HDD に入りません。

そこで元データを局面から棋譜に変換、着手分布は上位の 10 手だけにする、でデータを 46GB 程度(テキスト)にしてようやく取り扱えるサイズになりました。ちなみに LZ の Root で展開した手の数(着手分布の数)は平均 121 手程度あり AobaZero の 10 手程度と比較してかなり多いです。

また LZ の学習データには 9 路盤のものや、正しくないデータも含まれており、オープンソースで正しい棋譜を集める難しさが分かります。