

AobaZero の 2024 年のアピール文書

山下 宏

yss@bd.mbn.or.jp

1 AlphaZero の追試が最初の目的

AobaZero は Bonanza、LeelaZero のコードをベースに AlphaZero の追試をするべく MCTS + ディープラーニング で実装されてます。ネットワークは 3x3 のフィルタが 256 個の 20 block の ResNet でパラメータの個数は 2340 万個。棋譜生成をユーザの皆様と協力して行う分散強化学習です。オープンソースです*1。

2 AlphaZero の追試は 2021 年 4 月に終了

AlphaZero の将棋の追試は、2019 年 3 月から開始し、2021 年 4 月に 3900 万棋譜を作成して終了しました*2。2024 年 3 月 29 日現在、6746 万棋譜を作成しています。

3 追試終了後から +260 ELO、去年の選手権から +30 ELO

追試終了後からは +260 ELO、去年の選手権からだ +30 ELO ほど強くなっています。追試終了時では AlphaZero より +150 ELO 弱い、という推定でしたが現在は +260 なので AlphaZero を +110 ELO 程度、超えた棋力かもしれません。

4 去年の選手権からの主な変更点

- 学習棋譜の平均 playout 数を 1600 から 3200 に変更
- 先手勝ち局面の学習確率を減らす
- NVIDIA のドライバ更新時のエラー修正

4.1 学習棋譜の平均 playout 数を 1600 から 3200 に

ほぼ重みの棋力が停滞していたので playout 数を倍にしてより強い棋譜を生成するようにしました。当然生成時間は倍かかるので、生成棋譜数も 11000 棋譜/日から 9700 棋譜/日に下がってます。・・・ほとんど下がっていませんね。実は倍に変更してからかなりの資源を投入して棋譜生成に協力してくださった方がいたためです。この場をお借りして感謝いたします。棋譜の棋力はほぼ同じ条件で floodgate で走らせた限りでは平均 1600 の 3630 から平均 3200 で 3740 と +110 ELO 程度向上しているようです。この平均、とは 400playout ごとに Root の着手の訪問数の分布に変

化がないなら打ち切る、という手法です。最小 400、最大 12800 です。この KL 情報量を利用した探索の打ち切りは LeelaChessZero で使われています*3。この探索量の調整で 1 手 3200 固定より +100 ほど強くなります。

4.2 先手勝ち局面の学習確率を減らす

playout 数を倍にして棋力を上げたのですが、生成棋譜の先手勝率が 0.724 と 7 割を超えていました。PUCT の性質で勝率が悪い局面だと多くの候補手を調べて、さらに弱くなっていくケースがありそうです。そこで先手が勝った局面の学習確率を減らして初期局面の勝率が 5 割になる重みを再学習で作ってみたところ、棋力も +14 ELO 程度強いものが出来たので、これに差し替えました。その結果、先手勝率は 0.632 と 1 割近く下がりました。形勢判断としては正しくないですが棋譜生成においてはこちらの方が望ましいかもしれません。

4.3 NVIDIA のドライバ更新時のエラー

NVIDIA のドライバを最新にすると AobaZero が起動時に動かなくなる、という現象が発生しました。これは初回起動時に OpenCL の最適な設定を見つける行列計算を行う際に、OpenCL を動的にコンパイルしていたのですがコンパイラのエラーチェックが厳しくなったせいでした。芝さんなどの指摘で気づきました。ありがとうございます。

5 dlshogi 互換モデルで出場予定

AobaZero の棋譜で dlshogi のモデルを作ってみました。AobaZero と同じ 20 ブロック、256 フィルタの ResNet、Swish だと学習速度は 5.7 倍！探索速度の NPS は 3 倍ほど速いです。学習は 3090 だとミニバッチ 256 が最大だったのですがミニバッチ 2048 もあっさり動いたのにビックリでした。

1 手 1playout と 1 手 100playout、nps(RTX 3090) は表 1 です。ミニバッチはすべて 1 です (nps 測定では AobaZero が mb=17、dlshogi が mb=256)。dlshogi dr2 には Policy の強さ (1p)、探索 (100p) でもまだはつきり負けています。重みサイズが 2 分の 1 程度にも関わらず。384x30b にしてもまだ Policy の精度で負けてるので何か学習が正しくない気はします。ただ実時間だと nps が 3 倍近いのもあり (3090

*1 <https://github.com/kobanium/aobazero>

*2 <https://github.com/kobanium/aobazero/issues/54>

*3 <https://lczero.org/dev/wiki/>

[technical-explanation-of-leela-chess-zero/](https://lczero.org/dev/wiki/technical-explanation-of-leela-chess-zero/)

で AobaZero は 6000 程度) +40 ELO 程度 dlshogi モデルが強いので、これで選手権は参加する予定です。AobaZero は TensorRT に対応してないのと、複数 GPU での性能がいまいちなのもあって。計測は 24 手までの互角局面集利用で先後入れ替えて 800 局からです。

5.1 局面選択は Value に効果的？

学習は通常の AobaZero の学習と同じように学習される局面を選択しています*4。

先手勝率を 5 割になるように先手勝ち局面の選択割合を減らす、序盤の学習確率を減らす、勝率が 80 % 程度の手が 8 倍程度で学習されやすく、など。全局面の 50 % ほどが選択されるぐらいで。それを csa 形式に変換した後、aoba_to_hcpe3.py を改造して hcpe3 形式に変換しています。この選択をなしで hcpe 形式 (着手のみ) で学習させた場合は Policy の精度はいいのですが 1 手 100playout で 251 差、とはっきり弱く、局面の選択をするのは Value の精度に効いてるかもしれません。なお重複局面を平均化して削除する dlshogi の学習機能はどちらも有効にしています。

学習は 4000 万から 6690 万までの 2690 万棋譜から 5 億 7000 万局面を抽出してミニバッチ 2048 で学習率 0.1 から 0.001 まで Cosine Annealing で学習させました。さらに 0.00001 ぐらいまで下げた方がもう少し性能が上がるようです。

表 1 AobaZero(w4357) の dlshogi モデルに対する ELO

	1p/手	100p/手	nps
dlshogi dr2 (224x15b)	-61	-100	25300
AobaZero 256x20b	186	100	20500
AobaZero 256x20b(選択なし)	157	251	20500
AobaZero 384x30b	61	-58	5500

6 定跡は floodgate の AobaZero の棋譜から

定跡は floodgate で走らせている AobaZero の棋譜の局面を 30 万局面ほど探索させて作成する予定です。

7 対振り飛車の弱さが課題

floodgate で HoneyWaffle との棋譜を見ると相手が飛車を振っただけで評価値が +400 (勝率 65 %) と大きく出ます。学習棋譜でも振り飛車の割合は 2 % しかなく、ミレニアムや振り飛車穴熊、といった困いの知識もないので多様性に欠けて

る感じはします。この影響もあり特に対振り飛車は弱い感じですか。

8 人間の知識は使っていない、をおそらく継続

利きの情報の追加や 3 手詰、dfpn 詰などで AlphaZero からは離れてきましたが、まだ全体としては「人間の知識は使っていない」を継続していると考えています。

9 棋力の推移

図 1 が ELO の推移です。floodgate での測定レートの方が若干高めなのは Kristallweizen での棋力測定に用いている互角局面集 (24 手まで) には AobaZero が指さない穴熊や振り飛車が多数含まれているせいです。局面集を使わずに初手から指させた方が +100 ELO ほど強くなります。

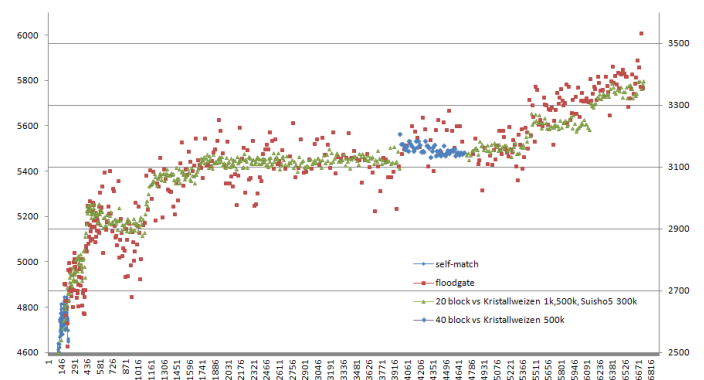


図 1 棋力の推移。右軸が floodgate のレート、横軸は棋譜数 (万)

10 5 年で 6700 万棋譜

6700 万棋譜、という膨大な棋譜を 5 年間で生成してきました。棋譜生成に協力していただいている皆様に感謝いたします。

*4 https://www.apply.computer-shogi.org/wcsc33/appeal/AobaZero/2023csa_appeal.pdf