

将棋の駒落ちの強化学習

山下 宏

yss@bd.mbn.or.jp

将棋の駒落ちを人間の知識なしでゼロから深層強化学習した。6種類の駒落ち(香落、角落、飛落、2枚落、4枚落、6枚落)と平手、の合計7種類を同時に学習し、勝率は5割になるように下手(先手)の強さを自動調節した。980万棋譜を自己対戦で作成した結果、平手から6枚落ちまで、上手、下手、どちらを持っても高段者レベルの着手を返すAIができた。

Reinforcement Learning for Shogi Handicap Games

Hiroshi Yamashita

We have developed deep reinforcement learning for Shogi handicap games without human knowledge. Handicaps are seven kinds. Lance(kyo ochi), Bishop(kaku ochi), Rook(hisha ochi), 2-Piece(ni-mai ochi), 4-Piece(yon-mai ochi), 6-Pieces(roku-mai ochi) and No handicap(hirate). Winrate are adjusted to keep 0.5 by weakening Black(shitate or sente) player strength. As a result of creating 9.8 million self-play games, AI could play reasonable moves at the level of a high dan player from No handicap to 6-Piece.

1 はじめに

AlphaZero^[1]は将棋のルールだけを教えて、機械がどのような将棋を指すのか、どこまで強くなるのかを明らかにした。本論文はそれを駒落ちで試す。新定跡を発見できるか、または2枚落ちの銀多伝、二歩突き切り、といった定跡を再発見できるか、が目標である。AlphaZeroとの大きな違いは初期局面が7種類あること、強さが違うプレイヤーの自己対戦で強化学習を行う点である。この実験「Aoba 駒落ち」はオープンソースで誰でも棋譜生成に協力可能である。棋譜や棋力の推移などはこちら^{*1}で公開している。

2 棋力の推移

図1は平手における棋力の推移である。自己対戦での1手1シミュレーション(左軸)、1手100シミュレーション対Kristallweizen^{*2}の1手40kノード(右軸)、および1手800シミュレーションのfloodgateでの実測(右軸)である。実測以外は互角局面利用^{*3}で400局の対戦結果から。学習率は0.001から開始して371万棋譜、573万棋譜でそれぞれ10分の1にした。後1回下げれば+50 ELO程度は強くなると思われる。

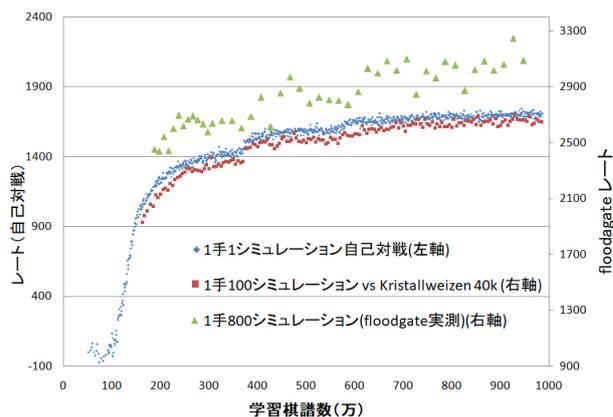


図1 強化学習のレーティングの推移

3 なぜ強さを調整するのか?

AobaZero^{*4}同士で6枚落ちを戦わせると、序盤から歩をただであさり取らせる、といった手を指す。これは駒落ちを学習させてないせいもあるが、対局開始直後から下手の勝率が99%になって歩を取る手も取らない手も、ほとんど同じ勝率になってしまうためである。これは上手も同じである。高すぎる、低すぎる勝率ではMCTSは2つの指し手の違いを認識できない。こういう性質のまま学習させるのは失敗する可能性が高いと思われる。

*1 <http://www.yss-aya.com/komaochi/>

*2 <https://github.com/Tama4649/Kristallweizen> 探索部は6.00

*3 <https://yaneuraou.yaneu.com/2016/08/24/>

*4 保木、山下、小林によるAlphaZeroの追試。2021年4月に追試を終了した。 <https://github.com/kobanium/aobazero>

4 強さの調整

強さの調整はいくつか方法が考えられる。

1. シミュレーション回数を1手800回から400回、200回など減らす。
2. 温度で調整する。
3. ある確率で合法手の中からランダムに選んだ手を指す。

図2は(1)の手法で、AobaZeroの重みw3020*⁵を用いて、1手のシミュレーション回数を変えた時のレート差である。

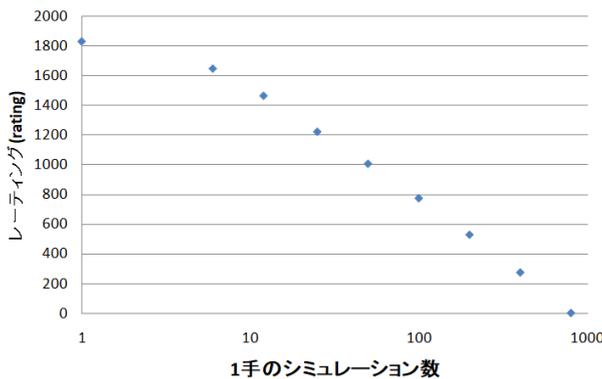


図2 1手のシミュレーション回数を変えた時のレート差 (対1手800シミュレーション)

1手1シミュレーションは1手800シミュレーションより1829 ELO 弱い。

(2)はAlphaZeroで棋譜にランダム性を持たせるために使われている手法で、ある手aを選ぶ確率は以下の式となる。

$$\frac{N(a)^{1/\tau}}{\sum_{i=1}^n N(i)^{1/\tau}}$$

$N(i)$ はMCTSでのRootにおける個々の手の訪問回数、 τ が温度で、 $\tau \rightarrow 0$ で常に最大訪問回数の手を、 $\tau = 1$ で訪問回数の割合で、 $\tau \rightarrow \infty$ ですべての手を同じ確率で選ぶ。

図3は温度を変えた時のレート差である。^{*6}この方法で調整する利点は実際に指した手でなく着手の分布を学習しているため、Policyの学習には悪影響が少ない点である。

面白いことに、(1)の1手1シミュレーションは1829 ELO弱くなるはずだが、これと互角になるのは(2)の温度1.3、レート差692 ELOの時であった。自己対戦でレート差が過大になる弊害である。レート差の最大は(2)の方が大きく、また学習の悪影響も少ないので(2)を採用した。実際は(2)と(3)を併用して強さを0から2557 ELO^{*7}まで変更可

^{*5} 先頭にwで3020番目に作成された重みを意味する。
^{*6} AobaZeroの自己対戦で、1手800シミュレーション(floodgateで3100点)から計算。温度が高いほど弱くなる。
^{*7} 2557 ELO 弱い、だと常に全合法手から乱数で1手選ぶ弱さになる。

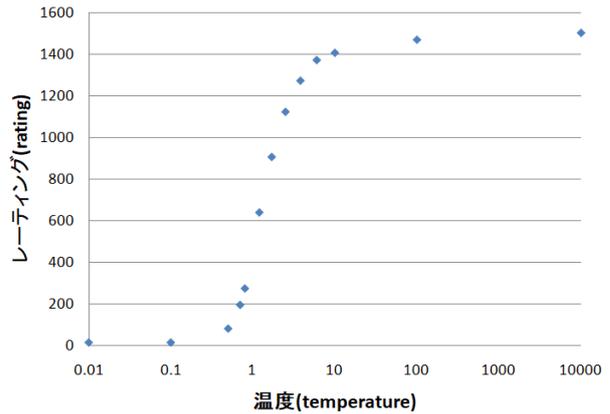


図3 温度を変えたときのレート差

能にし、0から1400までは温度で、それ以上は(3)としていた。ただ最大でも810 ELO程度のハンデしか生じなかったため用いたのは温度での調節のみである。

香落ちのレート調整は、最後に調整してから8000局の下手の勝率が0.53(+20 ELO)なら、その半分の10 ELOで下手を弱くする。他の手合いも同様である。弱くするのは下手(先手)のみで、上手(後手)は弱くしない。

図4はハンデレートの推移である。どの手合いのハンデも単調増加だが、香落ちだけ320万棋譜で下がっている。これはちょうど香落ちで上手が5手目△32飛と振飛車を指し始めたのと一致する。図5は下手の勝率の推移で全ての手合いでほぼ勝率5割となっている。

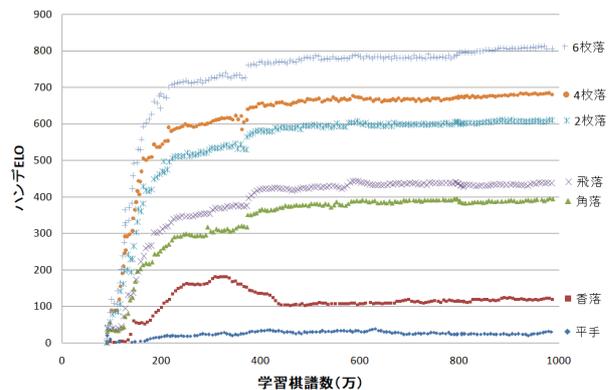


図4 下手のハンデレートの推移

5 将棋クエストのハンデレート差との比較

表1は将棋クエスト^{*8}の駒落ちで互角になるレート差とAoba駒落ちのレート差である。上手のレートが1300以下、1700以上と2つに分けて計算されている(1300~1700の人は内分)。将棋クエストは10分切れ負け用の設定。

^{*8} 棚瀬率による対局将棋サーバ <http://wars.fm/ja>

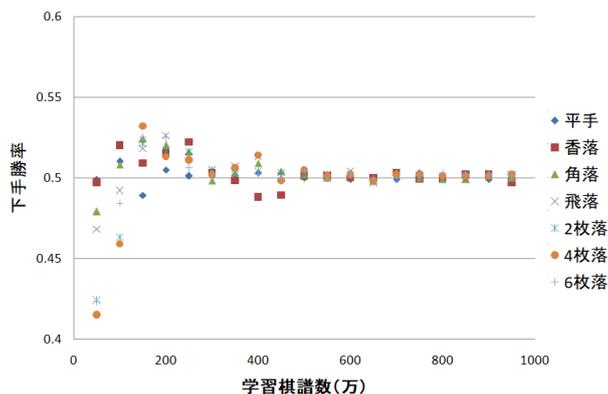


図5 50万棋譜ごとの下手(先手)勝率

表1 駒落ちハンデレート差。将棋クエストと Aoba 駒落ち

手合い	1300(5級)	1700(初段)	Aoba 駒落ち
平手			29
香落			119
角落	94	130	395
飛落	68	168	438
飛香落	106	181	
2枚落	225	299	610
4枚落			681
6枚落			809

Aoba 駒落ちと比べてハンデ差は小さい。飛落ちでも 200 以下で、飛落ちや角落ちは一般のアマチュアにとっては非常に小さい差のようである。また 1700 より 1300 の方がすべてハンデ差が小さく、棋力が低いと駒落ちのハンデを有効に活用できないことが分かる。

6 AlphaZero との違い

- 香、角、飛、2枚、4枚、6枚、平手、の7種類を同時に学習
- 勝率が5割になるように下手(先手)の棋力を弱く
- NNの入力に駒ごとの利き、利きの合計数、王手かどうかも追加。
- Policy Head を 11259 通りでなく 2187 通りに (dlshogi と同じ)。
- 投了なし
- 持将棋は 27 点法で。先手(下手)は 28 点、後手(上手)は 27 点で宣言。上手の落とした駒は上手の点数として計算^{*9}。奨励会の香落ちは 24 点法で、上手は 23 点あれ

*9 日本将棋連盟のルール解説 <https://www.shogi.or.jp/faq/rules/>

ば引き分け(落とした香車を足すと 24 点)。

7 他の AlphaZero と共通の主な設定

- 513 手目を指せれば引き分け
- Root の手にはディリクレノイズを常に追加
- 30 手目までは温度 1 で着手を選択。31 手目以降は基本は最大回数の手を選択。
- 31 手目以降は下手(先手)のみ、現在の調整 ELO に対応した温度で選択
- 30 手目までで現在の調整 ELO が 443 以上(温度 1 以上)なら下手のみ、その温度で選択
- 学習の損失関数は AlphaZero と同様、方策(Policy)分布の交差エントロピー、価値(Value)の二乗誤差、重みの L2 正則化項から
- 詰将棋はなし
- 探索木の再利用あり
- 1 手 800 シミュレーション固定。MCTS で探索。UCB のパラメータは AlphaZero と同じ $cBASE = 19652$, $cINIT = 1.25$, 勝率の初期値は負け。
- 探索部は AobaZero と同じ。探索内部で千日手、持将棋宣言判定あり。
- 直近の 50 万棋譜の全局面から 128 局面を乱数で選びミニバッチを作成して 1 回学習。
- 10000 回学習するごとに重みを作成。1 棋譜平均 128 手なら、1 局面を 1 回学習する割合。
- 10 分に一度、追加された棋譜数ぶんだけ Replay buffer を更新して学習。

8 学習の枠組み

ほぼ AlphaZero と同じである。クライアントは 7 分の 1 確率で各手合いの棋譜を生成し、棋譜収集サーバに送り、学習された重みを受け取る。学習部は棋譜を受け取り、作った重み、現在のハンデを棋譜収集サーバに送っている。

9 ニューラルネットワーク (NN) の入力

1. 盤上の駒の配置、 14×2 の 28 種類。
2. * 持駒の数、 7×2 で 14。最大数(歩なら 18)で割る。
3. * 過去に同一局面になった回数。1,2,3 回。
4. (1),(2),(3) の合計 $45(28 + 14 + 3)$ を現在と過去 5 手前まで、 $45 \times 6 = 270$ 種類。
5. 盤上の駒が動ける位置。 14×2 の 28 種類。現在局面のみ。
6. 利きの合計数。0,1,2,3,4 以上で 5 種類。を先手、後手で

10 種類。

7. * 王手か。1 種類。現在局面のみ。

8. * 現在手数。512 で割る。1 種類。

9. * 手番。1 種類。

10. * 手合い割。平手、香、角、飛、2 枚、4 枚、6 枚、で 7 種類。

* がついている項目は 9×9 が全部同じ値。持駒と手数以外は 0 か 1。合計で $318 \text{ 種類} \times 9 \times 9 = 25758$ 個の float。

10 NN の出力

着手の確率、 $27 \times 9 \times 9 = 2187$ 通り。勝率、1 通り。

11 NN の構造

AlphaZero と同じ、 3×3 のフィルターが 256 個、ResNet が 20 ブロック。

12 駒落ち初期局面集

下手、上手の棋力測定には学習棋譜の 16 手目の局面を出現回数が多い順にソートして上位 800 棋譜を開始局面とした。駒の配置もすべて 8 か所以上異なるように選んでいる。平手の場合ほとんどが相掛かりや角換わりの似た形と偏っており、初期局面集としては人間の棋譜から抽出した方が優れている。以降はすべて 400 対戦の結果である。800 棋譜の中からランダムで 400 棋譜を選んでいる。平手のみ、200 棋譜を選んで先後を入れ替えて 400 局としている。

13 Aoba 駒落ちの上手での棋力

測定には w870(920 万棋譜) を用いた。1 手 800 シミュレーションと互角になる水匠 4 改(下手)のノード数は 6 枚落 (10k)、4 枚落 (20k)、2 枚落 (30k)、飛落 (35k)、角落 (50k)、香落 (80k)、平手 (100k) であった。Bonanza 6.0 は 6 枚落 (30k)、4 枚落 (400k)、2 枚落 (1000k)、飛落 (10000k) であった。図 6 はそれを floodgate のレートに直したものである。水匠 4 改基準だと 6 枚落ちで 2580 と互角だが Bonanza 基準だと 2200 と互角である。これは水匠 4 改が平手では 10k で 2580 だが、6 枚落ちの下手ではもっと弱いためと思われる。Bonanza 基準の方が表 1 のレート差に近い。図 7 は水匠 4 改に対してシミュレーション数を 4 倍にした時のレートの伸びである。800 から 3200 の方が伸びが小さく、また 6 枚落ちに近い方ほど上昇も減っている。6 枚落ちの 1 手 3200 では勝ちの 84% が宣言勝ちであり、駒を落とすほど Aoba 駒落ちの宣言勝ちは多くなる。平手での強さは floodgate で 1 手 800 シミュレーションで 3000 点ほ

どである。同じ 20 ブロックの AobaZero が 3100 点なので 100 ELO 弱い。

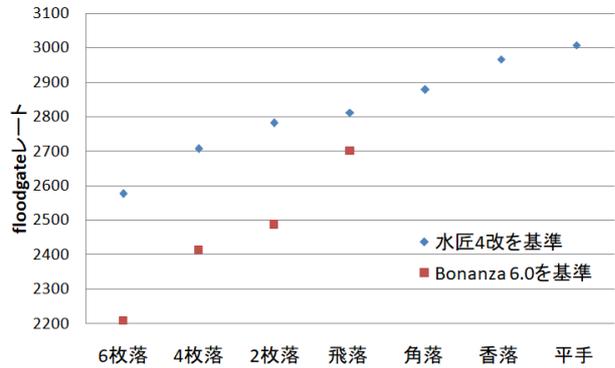


図 6 Aoba 駒落ち (1 手 800 シミュレーション) が上手で互角になる floodgate のレート

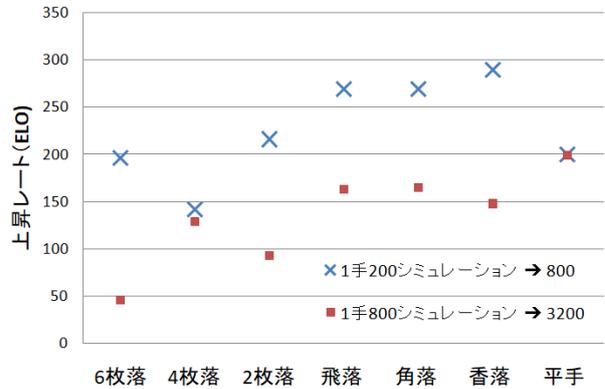


図 7 思考時間を 4 倍にしたときの Aoba 駒落ち (上手) の棋力の伸び。水匠 4 改の互角ノード数に対して

14 Aoba 駒落ちの下手での棋力

Aoba 駒落ちの 1 手 1 シミュレーションを下手にしたときの強さを調べた。水匠 4 改^{*10}、2 枚落ちの上手専用に王を中段に上げるように調整された水匠 U^{*11*12}、NNUE である水匠との比較のため KPPT 型 (3 駒関係に手番あり) として 2017 年選手権優勝の elmo^{*13}、深層学習を用いた dlshogi^{*14*15} を相手に選んだ。dlshogi 以外の思考エンジン

^{*10} 2021 年第 2 回世界将棋 AI 竜戦 TSEC の優勝プログラム。
https://twitter.com/tayayan_ts/status/1345619097556381697
 水匠 4 改 (2021 年 8 月 26 日)、水匠 U (2019 年 7 月 17 日)

^{*11} 2020 年 5 月の CSA 例会より

^{*12} 将棋倶楽部 24 で 2600 程度の相手に 2 枚落ちで互角

^{*13} https://mk-takizawa.github.io/elmo/howtouse_elmo.html

^{*14} 2021 年 8 月 15 日の竜戦長時間マッチで水匠に 2 勝 1 敗で勝つ。15 ブロックの ResNet を用いている。

^{*15} 第 2 回世界将棋 AI 竜戦エキシビジョン版 (2021 年 8 月 16 日)
<https://tadaoyamaoka.hatenablog.com/entry/2021/08/17/000710>

はやねうら王 6.50^{*16} を用いた。6.50 から駒落ちの宣言勝ち (Aoba 駒落ちと同じく、下手は 28 点、上手は 27 点で勝ち) にも対応しており、平手以外では「setoption EnteringKingRule value CSARule27H」を指定した。

図 8 は思考時間を 10 倍にしたときの上手の棋力の伸びである。Aoba 駒落ち以外は 6 枚、4 枚、2 枚、飛、角落ちで 10 倍で +100 ELO 程度の向上と小さい。表 2 の数値からも 4 枚、6 枚落ちでは下手が 1 手 1 シミュレーションと最小の弱さにもかかわらず、上手の勝率はきわめて低い。



図 8 思考時間を 10 倍にしたときの上手の棋力の伸び

15 重複なしの (ユニークな) 局面の割合

図 9 は 50 万棋譜ごとの 30 手目で重複なしの (ユニークな) 局面の割合である。手合いごとに傾向がはっきり違う。2 枚、角落ち、6 枚はほぼ全部ばらばらの局面にばらけている。平手は最新では 62% 程度である。乱暴に推測すると、2 枚落ちや角落ちは変化が広く、上手、下手、ともに何をやっても大差なく、香落ちには振飛車 (三間限定?) にするしか対策がなく、平手も相掛りか角換わり系統しか互角に戦えない、と判断しているのかもしれない。

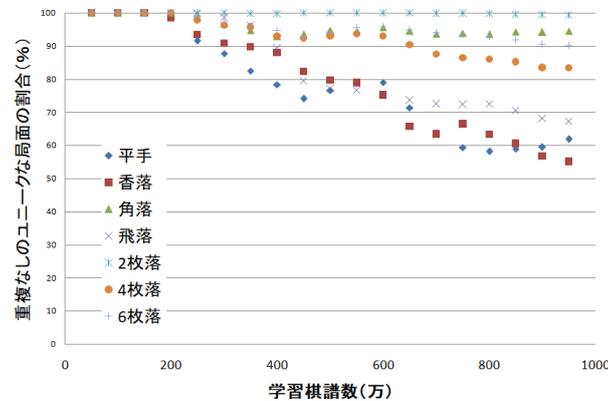


図 9 30 手目で重複なしの (ユニークな) 局面の割合

*16 <https://github.com/yaneurao/YaneuraOu/releases/tag/v6.50>

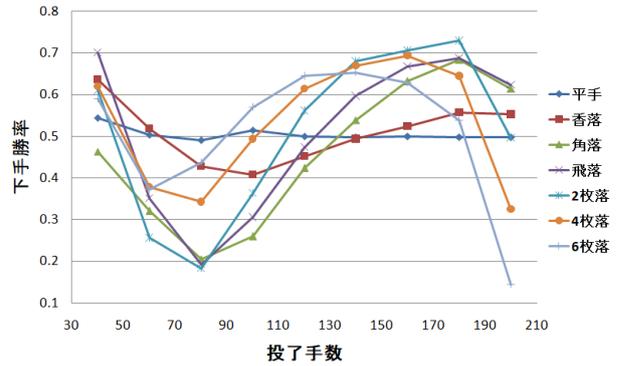


図 10 投了手数と下手勝率

16 投了手数と下手勝率

図 10 は投了手数と下手勝率である。予想ではミスをする下手は手数が多いほど勝率が下がる、であったが、80 手で勝率が下がり 160 手は上がる、逆の変化になっている。

17 カルバックライブラー情報量による探索打ち切り

794 万棋譜からカルバックライブラー情報量による探索の打ち切りを導入した。これはチェスの LCZero で採用された手法^{*17}で、通常 1 手に 800 シミュレーションを毎回行うが、100 シミュレーションした時の着手回数分布と 200 シミュレーションでの分布がほとんど同じなら探索を打ち切る、という手法である。最大 3200 シミュレーションまで 100 追加ごとに判定する。対 Kristallweizen では、ほぼ同じ探索時間で +71 Elo の効果があった。学習での結果は +26 ELO 程度の上昇^{*18}、と非常に小さく判断が難しい。

18 計算資源

学習には RTX 3090, 棋譜生成は RTX 3090, 2080Ti, GTX 1080 2 枚, 1050Ti, および Google Colaboratory の Pro (1072 円/月) を最大 18 アカウント使った。Colab Pro は GPU をブラウザ上で利用できる。24 時間経つと切断され再度クリックする必要がある。60 日間の合計で V100 が 32 回、P100 が 322 回、T4 が 478 回割り当てられた。K80 や P4 の場合はリセットした。開始から 4 か月で 1000 万棋譜近くを生成した。1 時間あたりの棋譜の生成速度は表 3 である (平均手数 122 手の場合)。

*17 <https://medium.com/@veedrac/leela-chess-test40-test50-and-beyond-c15896becfac>

*18 過去 50 万棋譜での重みの強さの平均比較。1 手平均 819 シミュレーションと若干増加した。kldgain-per-node = 0.000002

表2 Aoba 駒落ち (下手) に対する上手の勝率 (ELO) と思考時間 10 倍でのレート伸び

	水匠 4 改	伸び	水匠 U	伸び	elmo	伸び	dlshogi	伸び	Aoba 駒落ち	伸び
6 枚落ち (1 手 100k)	-880		-619		-655		-821		-427	
6 枚落ち (1 手 1000k)	-687	(+193)	-576	(+43)	-557	(+98)	-742	(+79)	-5	(+422)
6 枚落ち (1 手 10000k)	-596	(+91)	-530	(+46)	-481	(+76)	-969	(-227)	193	(+198)
4 枚落ち (1 手 100k)	-446		-260		-376		-569		-187	
4 枚落ち (1 手 1000k)	-352	(+94)	-256	(+4)	-228	(+148)	-498	(+71)	183	(+370)
2 枚落ち (1 手 100k)	-219		-74		-200		-439		-130	
2 枚落ち (1 手 1000k)	-143	(+76)	-26	(+48)	-85	(+115)	-350	(+89)	244	(+374)
飛落ち (1 手 100k)	294		249		216		63		60	
飛落ち (1 手 1000k)	381	(+87)	530	(+281)	346	(+130)	208	(+145)	485	(+425)
角落ち (1 手 100k)	386		384		254		242		134	
角落ち (1 手 1000k)	520	(+134)	474	(+90)	436	(+182)	369	(+127)	506	(+372)
香落ち (1 手 100k)	-140		-199		-354		-162		-50	
香落ち (1 手 1000k)	146	(+286)	123	(+322)	-12	(+342)	352	(+514)	576	(+626)
平手 (1 手 100k)	-12		-180		-226		99		0	
平手 (1 手 1000k)	360	(+372)	193	(+373)	135	(+361)	645	(+546)	481	(+481)

※ -880 は水匠 4 改の 1 手 100k の勝率が 0.006(-880 ELO) を意味する。

※ Aoba 駒落ち (下手) は 1 手 1 シミュレーション。香落ち、平手は 1 手 800 シミュレーション。

※ Aoba 駒落ち同士では下手の香落ち、平手は 1 手 10 シミュレーション。

※ 1 手 100k は 1 手に 10 万ノード (100,000) 探索の意味。

※ dlshogi の 1 手 100k、1000k、10000k は 1 手 1000、10000、100000 シミュレーション。

※ Aoba 駒落ちの 1 手 100k、1000k、10000k は 1 手 10、100、1000 シミュレーション。

表3 1 時間あたりの棋譜の生成速度

GPU	棋譜数	GPU	棋譜数
3090	736	1080	126
2080Ti	419	P4	82
V100	323	K80	46
P100	204	1050Ti	40
T4	148		

19 学習でみつけた指し方

もっとも印象深いのは香落ちで上手が振飛車を指すようになったことである。王を囲うのは 2 枚落ちまでで 4 枚落ち、6 枚落ちでは下手は王を囲わずに戦う。2, 4, 6 枚落ちでは上手は歩をただで取らせる手を気軽に指して長引かせようとする。発見した囲いは、中住まい、矢倉、美濃囲い、高美濃、右玉、金無双、雁木、などで穴熊は指さない。以降で紹介する棋譜は 1 手 800 シミュレーションで特定の重み同士をノイズなしで自己対戦させたもので下記^{*19}で閲覧できる。再現性がある。

*19 http://www.yss-aya.com/komauchi/no_noise/sample7.html

19.1 平手

相掛かり、角換わり腰掛銀、矢倉、脇システム、横歩取り、右玉などを指す。傾向としては AobaZero と似ている。ただ先手、後手ともに振飛車は指さない。

19.2 香落ち

唯一、香落ちのみ常に振飛車が出現する。香車がない左側に王を囲っても堅くないため、定跡では上手は振飛車がよいとされ、それを再発見できた。上手は常に三間飛車で、それに対して下手も居飛車でなく、三間の振飛車で対抗する。つまり相振飛車となっている。下手は石田流の形を好んで狙う。上手にしっかり囲われるのを嫌い、早めに乱戦になることが多い。図 11 は w454(504 万棋譜) の 50 手目で下手は美濃囲い、上手は銀冠の相振となっている。最近の AI では飛車を振るだけで勝率が 1 割以上上がるほど振飛車は不利とされ、選手権の決勝ではほぼ登場しない。しかし状況によっては振飛車も十分有効な戦法のようなのである。

19.3 角落ち

上手は居飛車で端攻めを警戒してか早めに△ 22 銀と上がる囲いが多い。下手は居飛車だけでなく、三間飛車や向かい飛車にも組む。図 12 は w929(979 万棋譜) で下手の三間飛車に 1 筋を詰められた上手が端から逆襲している。このように上手が先に端攻めを狙う展開は多い。

19.4 飛落ち

上手は初手△44歩から角交換を拒否しようとするが、下手も2手目▲46歩から角交換を強要する。最新のw935(985万棋譜)では図13のように上手が角道を突かずに角交換を避ける形が出現した。早い段階での角交換は下手有利(上手は打ち込みの隙が多い)、を理解しているようである。

19.5 2枚落ち

二枚落ちには「二歩突き切り」と「銀多伝」という古典定跡^[2]が存在する。この2つを発見できるか、はこの実験の一つの目的であった。結論から言えばどちらも発見できなかったが「二歩突き切り」に似た形は出現した。図14はw850(900万棋譜)の12手目である。後1手▲35歩と指せば二歩突き切りとなる。4筋の位は早々に取るのだが▲35歩と3筋の位を取る形はほぼ指さない。

図15はw871(921万棋譜)の39手目で下手はこの形によく組む。雰囲気は銀多伝に似ている。違いは銀が2枚縦に並ぶ代わりに▲46銀▲47金と並べること、王は▲48玉(銀多伝は▲38玉)のままで、78の地点は銀で守ることである。

19.6 4枚落ち

上手は飛車を成られるのを極端に嫌がり、歩をただで捨てては手を遅らせようとする。図16はw940(990万棋譜)の19手目で△84歩とただの場所に歩を突いている。これを▲同角は△83金▲66角△94香で端攻めが遅くなる。

19.7 6枚落ち

図17はw939(989万棋譜)の3手目で△24歩が相手の飛車先の歩を突く斬新だが、よく出現する手である。下手は1筋、もしくは9筋を飛、角、桂、だけで破ろうとする。上手はただで歩を取らせて右金を進出する手を稼ぐ形が多い。



図12 角落ち、w929(979万棋譜)の25手目。下手が三間飛車



図13 飛落ち、角道を突かず角交換を拒否。w935(985万棋譜)



図11 香落ちでは相振飛車になる。w454(504万棋譜)の50手目

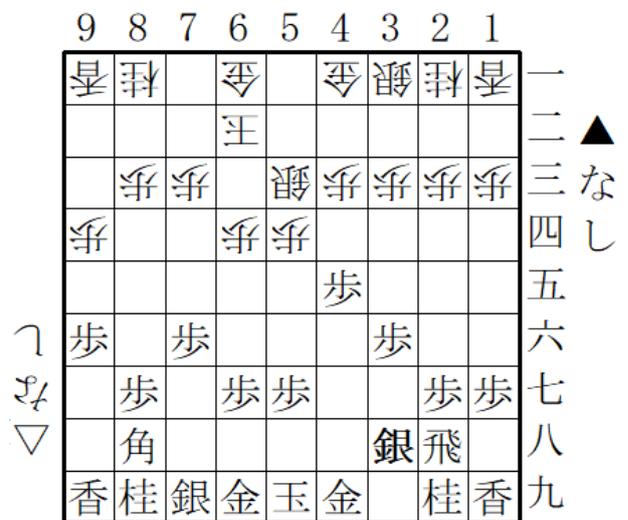


図14 2枚落ち、二歩突き切りに似た形。w850(900万棋譜)



図 15 2 枚落ち、銀多伝に似た形。w871(921 万棋譜)



図 16 4 枚落ち、ただの場所に歩を突く。w940(990 万棋譜)



図 17 6 枚落ち、△ 24 歩は挑発的。w939(989 万棋譜)

20 930 万棋譜から 980 万棋譜でもっとも多く出現した 16 手目までの手順

平手 ▲ 2 六歩△ 3 二金▲ 7 八金△ 8 四歩▲ 2 五歩△ 8 五歩▲ 3 八銀△ 7 二銀▲ 9 六歩△ 8 六歩▲ 同 歩△ 同 飛▲ 8 七歩△ 8 二飛▲ 3 六歩△ 1 四歩

香落 △ 3 四歩▲ 1 六歩△ 3 二飛▲ 7 六歩△ 4 四歩▲ 7 八飛△ 3 五歩▲ 7 五歩△ 8 二銀▲ 7 四歩△ 6 二玉▲ 7 三歩成△ 同 銀▲ 7 四歩△ 8 二銀▲ 3 八銀

角落 △ 4 二玉▲ 7 八銀△ 5 二金右▲ 1 六歩△ 8 四歩▲ 7 六歩△ 5 四歩▲ 6 六歩△ 6 二銀▲ 3 六歩△ 8 五歩▲ 7 七角△ 3 二玉▲ 6 七銀△ 7 四歩▲ 5 六歩

飛落 △ 4 四歩▲ 4 六歩△ 3 二金▲ 3 八銀△ 5 四歩▲ 4 七銀△ 3 四歩▲ 5 六銀△ 4 二銀▲ 7 六歩△ 4 三銀▲ 4 五歩△ 同 歩▲ 2 二角成△ 同 金▲ 4 五銀

2 枚落 △ 5 四歩▲ 4 六歩△ 6 二銀▲ 3 八銀△ 4 四歩▲ 4 七銀△ 5 三銀▲ 4 八飛△ 4 二銀上▲ 7 六歩△ 4 三銀▲ 5 六銀△ 5 二金右▲ 4 五歩△ 同 歩▲ 同 飛

4 枚落 △ 6 四歩▲ 7 六歩△ 5 二金右▲ 1 六歩△ 2 二銀▲ 1 五歩△ 4 二玉▲ 1 八飛△ 6 三金▲ 1 四歩△ 同 歩▲ 同 飛△ 1 三歩▲ 1 八飛△ 3 二金▲ 1 七桂

6 枚落 △ 4 二玉▲ 1 六歩△ 2 二銀▲ 1 五歩△ 2 四歩▲ 2 六歩△ 3 二金▲ 7 六歩△ 4 四歩▲ 同 角△ 5 二金▲ 6 六角△ 8 二銀▲ 2 五歩△ 同 歩▲ 1 七桂

21 おわりに

強さの異なるプレイヤー同士の自己対戦で、駒落ちの学習が成功した。7つの初期局面をすべて互角と評価でき、強さの調整により、それぞれに適切なハンデレート差がついた。角落ち以上の上手、下手で既存のソフトよりスケールがよいAIが作れた。香落ちでは上手の振飛車を再発見できた。

22 謝辞

Aoba 駒落ちは AobaZero を改造して作成された。棋譜配布サーバや高速化をされた保木邦仁氏、貴重な指摘を頂いた小林祐樹氏、将棋クエストの駒落ちハンデを教えていただいた棚瀬寧氏、棋譜生成に協力して頂いた皆様、とくに毎夜、1500 棋譜/時 を提供していただいた方に感謝いたします。

参考文献

- [1] D. Silver, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, Science, 362, 1140-1144 (2018)
- [2] 【決定版】駒落ち定跡, 所司和晴, マイナビ出版 (2010)