

モンテカルロ木探索のパラメータ調整による 人間の指し手との一致率の向上

杵渕哲彦[†] 伊藤毅志[†]

概要：ゲーム AI はこれまで強さの向上を目的として発展してきたが、近年は人間の程よい対戦相手となるなど強さ以外の研究も行われており、その一つとして人間の選択結果を予測することが挙げられる。本研究では将棋を題材にあまり強くない人間の指し手を模倣することを目的とし、MCTS の探索傾向を変化させて目的の強さのプレイヤーが行うような探索に近づけることで、一致率の向上を試みた。既存の探索を全く行わない手法では予測出来ない指し手を予測できたが、既存手法と比較して一致率は下回った。

キーワード：コンピュータ将棋、人間の模倣

Improved Move-Matching Accuracy with Human Moves by Parameter Tuning for Monte Carlo Tree Search

TETSUHIKO KINEBUCHI[†] TAKESHI ITO[†]

Abstract: Game AI has been developed for the purpose of improving strength, but recently there has been some research done outside of strength, such as being just the right opponent. This research aim to imitate human player's move in Shogi. By changing the search tendency of MCTS to imitate the search conducted by human players, we attempted to improve move-matching accuracy. The proposed method was able to predict moves that could not be predicted by the existing method, but move-matching accuracy was lower than that of the existing method.

Keywords: Computer Shogi, Human Imitation

1. はじめに

ゲーム AI の研究は強さの向上を主な目的として様々な技術や手法が開発され、チェスや将棋、囲碁のような二人零和有限確定完全情報ゲームでは人間のトッププレイヤーを上回る強さを実現した。また近年では、不完全情報ゲームであるポーカーや麻雀、多人数ゲームであり他のプレイヤーとの協力が求められるFPSなどでも高いパフォーマンスを示している。

これまでのゲーム AI 研究で得られた局面の価値の推定手法や局面の探索手法などの技術を用いて、様々な強さのプレイヤーを相手に良い勝負を演じることが出来る接待プレイや、現在の状態が強さの異なる様々なプレイヤーにとってどの程度の難易度なのかの推定など、強さ以外を目的とした研究が行われてきている。これらの課題の一つとして、様々な強さの人間のプレイヤーの選択結果を予測することが挙げられる。

本研究で題材とする将棋では、指し手の選択は読みと大局観を使って行われる。ある局面で強いプレイヤーにとっては当たり前な指し手であっても、あまり強くないプレイヤーにとってはそれが選択できない場合は、この読みと大局観の精度の差に基づくことが多い。読みと大局観は将棋 AI で

は探索と評価に相当することから、様々な強さのプレイヤーの指し手を予測するには、探索と評価の精度を調整すればよいと考えられる。

そこで本研究では探索の精度の調整に着目する。既存手法として $\alpha\beta$ 探索で深さを浅くすることで AI を目標の棋力に調整する方法があるが、この手法では人間の指し手との一致率はあまり良くない。理由として、目標の棋力と同等の人間の読みと比較すると、人間の読みよりも探索の深さが浅く、人間よりも多くの候補手を考慮していることが多いなど、人間と AI の探索木の形状が大きく異なることが考えられる。このことから、人間の指し手を教師として AI の探索の精度を調整し、探索傾向を人間の読みに近づけることで、人間の指し手との一致率の向上を試みる。

2. 関連研究

特定の強さの人間の選択結果を模倣する研究として Maia[1]が挙げられる。Maia はチェスを題材に、AlphaZero のネットワーク構造を用いて、様々なレーティング帯の人間の棋譜を教師として、レーティング帯ごとにモデルを作成することで、既存手法よりも人間の指し手との一致率を向上させた手法である。この研究では着手予測時に探索を行うことで一致率が低下するという結果が報告されている。

[†] 電気通信大学大学院 情報理工学研究科

Maia の手法が将棋でも有効であり、局面によっては探索を行う既存の将棋 AI を用いることで一致率が向上することを示した研究として小川らの研究が挙げられる[2]。小川らは人によって着手が分かれる局面を予測モデルの確信度によって判定し、高レートのプレイヤーの棋譜で確信度が低い局面では、探索を行っている既存の強い将棋 AI を使うことで、Maia の手法よりも一致率を高められることを示した。一方で低レートのプレイヤーの棋譜では一致率の有意な向上は見られなかった。一致率が向上した理由としては、探索を行わない場合は数手後にならないと効果が分からない難しい手を当てるのが困難だからと予測している。この読まないと当てられない手という考えは低レートのプレイヤーの手を当てる場合にも適用できるのではないかと考える。低レートのプレイヤーも読みは行っているが、高レートのプレイヤーと比較すると考慮できる手の数は少なく深さも浅く、読み抜けなども発生する不完全なものだと思われる。この不完全な探索を再現することで、読みが必要な低レートのプレイヤーの手が当てられるようになり、一致率が向上させられるのではないかと考えられる。

3. 提案手法

3.1 概要

提案手法では、DeepLearning 系の将棋 AI の探索で用いられるモンテカルロ木探索(MCTS)を対象に、プレイアウト数や方策の温度パラメータ、PUCT アルゴリズムの定数などを調整して探索傾向を変化させることで、人間の指し手との一致率の向上を試みる。

3.2 APV-MCTS

APV-MCTS は方策ネットワークと価値ネットワークを用いて探索するモンテカルロ木探索の手法であり、各ノードでは(1)の式が最大となる手を選択する。 s_t は現在の局面、 a は局面 s_t で選択できる候補手を示す。 $Q(s_t, a)$ は局面 s_t における候補手 a の価値を表しており、UCB1 の式での期待値項に当たる。また $U(s_t, a)$ は探索回数が少ない手を多い手よりも優先させるための値で、UCB1 の式でのボーナス項に当たる。

$U(s, a)$ は(2)の式で求められ、 $P(s, a)$ は方策ネットワークで予測した着手確率、 $N(s, a)$ は局面 s における候補手 a の訪問回数を表す。したがって、 $U(s, a)$ は方策ネットワークによる着手確率が高い手を優先し、まだ訪問回数が少ない手を優先することを表しており、 c_{puct} の値が大きいほど優先度が大きくなることを意味する。

$$Q(s_t, a) + U(s_t, a) \quad (1)$$

$$U(s, a) = c_{puct} P(s, a) \frac{\sqrt{\sum_b N(s, b)}}{1 + N(s, a)} \quad (2)$$

4. 実験

4.1 目的

実験では低いレーティング帯の人間の棋譜で学習したモデルを用いて、Policy の値が最大の指し手を選ぶ Maia で採られた手法と、 c_{puct} の値と方策の温度パラメータを変化させることで探索傾向を変化させた MCTS を用いる提案手法で、実際に人間が指した手との一致率を比較する。

4.2 実験方法

将棋クエストのレーティング 1300~1399 の棋譜を用意し、32 万局を教師データ、100 局をテストデータとした。また、序盤は定跡により似た局面が出やすく学習に悪影響を及ぼす可能性があるため、40 手までの局面は除外した。本実験では DeepLearning を用いて方策ネットワークと価値ネットワークの学習を行う山岡[3]の python-dlshogi2 を用い、現局面を入力としその局面で指された指し手を予測するよう学習させた。

次に実験時の MCTS のプレイアウト数を定めるため、python-dlshogi2 のデフォルト値である $c_{puct} = 1.0$ 、温度パラメータ $= 1.0$ の条件でプレイアウト数 n を変化させて一致率を測定したところ、 $n = 400$ で最大となった。

この結果からプレイアウト数を 400 とし、 c_{puct} の値と方策の温度パラメータを変化させて一致率を測定した。また、Policy の値が最大の指し手を選ぶ手法で予測できる手と、提案手法で予測できる手に差異があるかの調査も行った。

4.3 結果

Policy の値が最大の指し手を選んだ場合の一致率は 45.35%だった。提案手法の結果を表 1 に示す。いずれの条件でも Policy の値が最大の指し手を選ぶ場合の一致率を下回った。

表 1 c_{puct} と温度パラメータを変動させた際の一致率

温度/ c_{puct}	0.1	1.0	10
0.25	40.9%	41.3%	41.0%
0.5	39.5%	41.4%	41.5%
1.0	37.2%	40.1%	41.8%
2.0	28.5%	34.3%	41.9%
4.0	22.2%	26.2%	39.34%

提案手法の中で最も一致率が高かった $c_{puct} = 10$ 、温度パラメータ = 2.0 の条件でのみ一致した局面と、Policy の値が最大の指し手を選んだ場合でのみ一致した局面の結果を表 2 に示す。

表 2 Policy の値最大と提案手法で一致した指し手の割合

	提案手法で一致	提案手法で不一致
Policy 最大で一致	38.8%	6.5%
Policy 最大で不一致	3.0%	51.6%

4.4 考察

Policy の値が最大の指し手を選んだ場合と比較して提案手法の一致率が下がった要因として、MCTS のパラメータを調整して探索傾向を調整する方法では低いレーティング帯のプレイヤーの探索傾向をうまく模倣することが出来なかったためと考えられる。一方でパラメータを調整した探索によって予測出来た指し手も確認されたことから、低いレーティングの指し手の予測では探索が有効に働く場合があると考えられる。

5. おわりに

本研究では、低いレーティングのプレイヤーの指し手を対象に、このレベルのプレイヤーが行う不完全な探索を模倣し、読まないと当てられない手を予測することで一致率の向上を試みた。結果としてパラメータを調整して探索傾向を変化させた MCTS ではうまく模倣できず、Policy の値が最大の手を選ぶ手法には一致率が及ばなかった。

今後の課題として、人間の探索の傾向や読むことが出来る候補手の数、深さなどの知見を参考に、低いレーティングのプレイヤーが行う不完全な探索を模倣できる手法を検討、開発していきたい。

参考文献

- [1] McIlroy-Young, R. and Sen, S., and Kleinberg, J., and Anderson, A., Aligning Superhuman AI with Human Behavior: Chess as a Model System, KDD, 2020
- [2] 小川竜欣, シュエ ジュウ シュエン, 池田心, 着手予測モデルが予測しづらい局面の考察・分類と確信度を利用した一致率の向上, ゲームプログラミングワークショップ 2022 論文集, 180-186, 2022
- [3] 山岡忠夫, 加納邦彦, 強い将棋ソフトの創りかた Python で実装するディープラーニング将棋 AI, マイナビ出版, 2021