

Title	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
Author(s)	小川, 竜欣; シュエ, ジュウシュエン; 池田, 心
Citation	ゲームプログラミングワークショップ2022論文集, 2022: 180-186
Issue Date	2022-11-04
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/18237
Rights	<p>社団法人 情報処理学会, 小川 竜欣, シュエ ジュウシュエン, 池田 心, ゲームプログラミングワークショップ2022論文集, 2022, 180-186. ここに掲載した著作物の利用に関する注意 本著作物の著作権は情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 Notice for the use of this material: The copyright of this material is retained by the Information Processing Society of Japan (IPJS). This material is published on this web site with the agreement of the author (s) and the IPJS. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPJS if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright (C) Information Processing Society of Japan.</p>
	第27回ゲームプログラミングワークショップ(GPW-



Description	22) , 2022-11, 電気通信大学
-------------	-----------------------

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

小川 竜欣^{1,a)} シュエ ジュウ シュエン^{1,b)} 池田 心^{1,c)}

概要: 最近では、強いゲーム AI 研究だけではなく、人間を指導したり、楽しませたりするゲーム AI 研究も行われている。人間の方策（局面ごとにどの着手をどの割合で選ぶか）を模倣できるモデルは、対局 AI、教師 AI や、詰将棋の難易度評価など、様々なことに応用が可能である。本研究では、局面から着手を予測する教師あり学習を行う際に、人間の着手予測においてチェスで効果的な予測モデルが、将棋でも有用であることをまず示した。その後、人間の着手をどれくらいの確率で予測していたか、という評価指標である尤度を使用して、予測モデルが予測しやすい局面や、予測しにくい局面について考察・分類を行った。また、予測モデルの確信度を用いて、確信度が低い状況で、dlshogi や水匠などの探索を利用している将棋エンジンを採用することで、一致率を高められることを示した。

キーワード: コンピュータ将棋, 人間中心の AI

Consideration and classification of positions that are difficult for move prediction models to predict

TATSUYOSHI OGAWA ^{1,a)} CHU-HSUAN HSUEH ^{1,b)} KOKOLO IKEDA ^{1,c)}

Abstract: Recently, researchers in the game AI field focus on not only strong AI players but also AI that teach or entertain human players. With models that can imitate human players' policies (the probability of each move for each position), we can apply these models to various tasks, such as opponent AI, teacher AI, and difficulty evaluators of Tsume-Shogi. In this research, we first employ supervised learning to create models that predict human moves from positions, an approach that has been applied to chess. Similar to the results obtained in chess, our prediction models for shogi can predict human moves well in general. We use likelihood, a measure of how well the models predicted the human moves, to consider and categorize the positions that are easy or difficult for the models to predict. Using the confidence of the prediction models, we also show that in situations where the confidence is low, the move matching rate can be increased by employing other programs that use search, such as dlshogi and Suisho.

Keywords: Computer Shogi, Human-centered AI

1. はじめに

将棋や囲碁、チェスといった二人完全情報確定零和ゲームにおいて、ゲーム AI は人間のトッププレイヤー以上に強くなった。最近では強いゲーム AI に関する研究だけではなく、人間を指導したり、楽しませたりするゲーム AI の

研究も行われている。楽しませるゲーム AI で重要な要素として、「不自然な着手の抑制」が挙げられる [1] ことを踏まえると、人間の方策を模倣できるゲーム AI の実現は重要な研究目標の一つだと考えている。

もし棋力帯ごとの、もしくは特定の人間プレイヤーの手を十分な精度で予測できれば、対局相手 AI として人間の着手を再現できるだけでなく、詰将棋の難易度を評価したり、教師 AI として手加減に向いている局面に誘導できたりする。

そこで着手予測の精度が問題になるが、人間の着手を模

¹ 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology,
Nomi, Ishikawa 923-1211, Japan

a) ogawa.tatsuyoshi@jaist.ac.jp

b) hsuehch@jaist.ac.jp

c) kokolo@jaist.ac.jp

倣したゲーム AI を評価する際は、指標として一致率がよく使用される [2][3][4]。一致率は、人間を模倣できているかの評価についてある程度有用であるが、人間とは一致していない着手については、人間から見て惜しい手だろうと全くありえない手だろうと同じ評価である、という弱点がある。例えば、一致率がある程度高いゲーム AI だったとしても、対局する際に人間から見て全くありえない手を指されれば、プレイヤーは不自然だと感じ楽しむことが難しくなる。

この一致率に代わる評価指標として、予測モデルが人間の指し手をどれくらいの確率で予測していたかという指標である、尤度を考える。人間を模倣するモデルの最終的な目標の一つは、人間の着手分布と等しい着手確率分布を再現することであり、尤度はこのときのみ最大になる。このように、尤度には一致率にはない優れた特徴を持っているが、モデルが着手確率を出力しなければ計算できないという欠点もある。

本研究では、予測モデルの尤度が高い局面・低い局面の考察・分類を行った後、予測モデルの確信度によって探索が必要な局面を判定し、人間をより模倣できるかを確認する。

2. 関連研究

Maia[2] は、AlphaZero[5] のネットワーク構造をもとにしているが、自己対局の代わりに人間の棋譜から学習を行う深層学習チェスプログラムである。チェスにおいては現在、人間の着手を予測する最も効果的な研究の一つである。この研究では、着手予測を行う際に、対局者の棋力帯によって学習するモデルを分けることで、人間の着手予測や、人間が次に大きなミスをするかという予測をより正確に行っている。この研究では着手予測に探索を用いると性能が悪化するという理由から、一切の探索を行っていない。しかし、Maiaと同じ学習モデルを使用しても、棋力に応じて適切なパラメータを設定して探索を行えば、より正確に着手を予測できることを示した研究がある [3]。ここから、探索を利用することで、より正確に着手予測できる状況が存在すると推測できる。

人間の指し手との一致率を高める将棋 AI 研究として、流れの表現を試みた研究がある [4]。この研究では、局面評価関数と探索に基づく手の予測器と、遷移確率関数に基づく手の予測器を結合し、ランキング学習を行うことで指し手の一致率を有意に高めている。

複数の将棋エンジンの局面評価値から 1 つの指し手を選択する合議アルゴリズムについての研究がある [6]。この研究では、合議アルゴリズムを Boosting 法のような手法の一種として捉えており、局面の優劣評価から最も高い評価値を返した将棋エンジンの指し手を選択することで、将棋エンジンを有意に強くすることに成功している。

3. 提案手法

【概要】本研究の概要としては、まず Maia のように AlphaZero をもとにしたネットワーク構造で将棋でも一致率を高めることができるか、そして学習する棋譜をレート帯ごとにわけること、そのレート帯の一致率を比較的高めることができるかを確認する。その後、予測モデルで尤度が高い局面、低い局面について考察・分類を行い、どのような局面を予測できればモデルの尤度・一致率を向上させられるかについて分析を行う。また、使用した予測モデルの推論の確信度を利用して、確信度が低い局面では探索を利用するモデルの着手予測を採用することで、一致率を向上させられると期待している。

【一致率】クラスが K 個ある多クラス分類に使用されるニューラルネットワークでは、入力を x 、出力を u_k とすると、 x がクラス C_k に属する確率 $p(C_k|x)$ は、

$$p(C_k|x) = \frac{\exp(u_k)}{\sum_{j=1}^K \exp(u_j)}$$

で表されることが多く、 $p(C_k|x)$ が最大となるクラスに分類される。将棋の場合では、局面 x について、合法手が K 個あるとすると、指し手 $C_k (k = 1 \sim K)$ の確率 $p(C_k|x)$ が出力され、 $k = 1 \sim K$ までの分布はそのまま方策関数 (policy function) として使用できる。一致率は、 $\arg \max_k p(C_k|x)$ が実際の指し手と一致しているかという評価指標であり、人間の着手の模倣を目指す研究の評価指標としてよく使用されている。しかし、人間の着手が分かれるような局面では、一致率ではうまく評価できない局面もある。例として、人間の着手 A が 60%、着手 B が 30%、着手 C が 10% で指されるような局面を考える。この場合、一致率で評価を行うと、モデルの着手確率分布が A : B : C = 90% : 5% : 5% でも、34% : 33% : 33% でも同じく一致率は 60% である。理想的な着手確率分布は 60% : 30% : 10% であることを考えると、この局面は一致率でうまく評価できていないといえる。

【尤度】尤度は、確率密度関数において確率変数に観測値を代入したもので、将棋においては、人間の指し手が C_{human} だった場合の $p(C_{human}|x)$ の値である。つまり、実際の指し手をどれくらいの確率で予測していたか、という評価指標である。複数の手について考えるとき、全体の尤度は各着手の尤度の積で表され、この値は、人間の着手分布とモデルの着手確率分布が等しいときのみ最大になる。よって、尤度は人間の着手の模倣を、より正確に評価できる指標だといえる。

【推論の確信度】将棋の着手予測において、多くの人間の着手が一致する局面と、人によって着手が分かれる局面では、モデルの着手確率分布の形は異なると予想している。また、一見人によって着手が分かれそうだが、探索を行え

ば多くの人間が最善手を選択できる局面も存在すると考えている。もしこのような局面を自動で判定することができれば、その局面にのみ探索を行うことで、着手の一致率や尤度を高めることができる。本研究では、このような局面の自動判定に推論の確信度 (confidence) を使用する。多クラス分類における $\max_k p(C_k|x)$ は分類したクラスの事後確率の予測値であるため、推論の確信度として解釈可能である [7]。推論の確信度は、何も対策を取らないと本来あるべき値よりも高くなる、すなわち「自信過剰」となる傾向があることが経験的に知られている [8]。「自信過剰」になっているモデルでは、予測が誤っているのに確信度が高いということが頻繁に起こり、 $\max_k p(C_k|x)$ をそのまま確信度として使用できないという問題が発生する。この問題については、実験 2 で検証を行う。

4. 実験 1：人間棋譜の学習

4.1 実験 1 の内容

実験 1 では、棋力ごとに棋譜を分けて、方策関数と価値関数の教師あり学習を行った。データは、将棋クエストの棋譜約 300 万局から条件を指定して抽出した棋譜を使用した。まず、時間切れ負けの棋譜を除外した。理由は、一番操作しやすい駒を動かすといった時間切れ負け特有の行動がありうるためである。また、対局者間のレート差が 50 以下の対局に絞った。これは、不利な局面でも最終的に高レート側が勝っている、といったレート差による結果の偏りを防ぐためである。これらの条件を満たした約 76 万棋譜を対局者の平均レートごとに分け、6 等分した。各データ群のレート帯は、以下のとおりである。

- グループ 0：R1433～R1591
- グループ 1：R1591～R1655
- グループ 2：R1655～R1708
- グループ 3：R1708～R1768
- グループ 4：R1768～R1855
- グループ 5：R1855～R2140

それぞれのデータ群の 90% を学習データ、5% を検証データとして学習を行い、残りの 5% をテストデータとして評価に使用した。また、手数が 50 手目以降の局面を使用した。序盤を除外した理由は、似た局面が多く学習に悪影響を及ぼす可能性があるためである。

方策ネットワークと価値ネットワークを 1 つのネットワークとして同時に学習するマルチタスク学習を行った。ネットワークの構造や学習オプションについては、山岡の書籍 [9] を参考にした (付録：ネットワークの構造)。山岡の書籍と大きく異なる部分は、入力に直近局面を含めた点である。

Maia の研究では、直近局面を入力に含めることで、着手予測の精度が有意に改善すると述べられている。本実験でも直近 12 局面を入力に含めると、現局面のみから着手

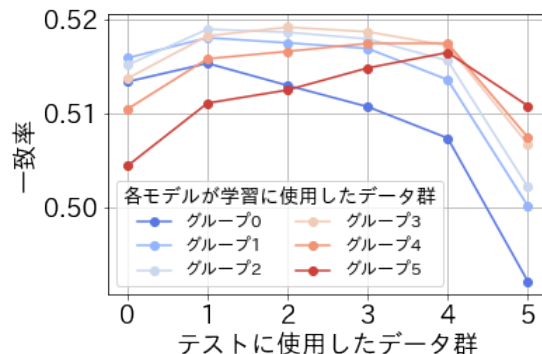


図 1 各モデルとテストに使用したデータ群との一致率の関係

予測するモデルと比較して一致率が改善したため、直近 12 局面を含めたモデルを採用した。

4.2 実験 1 の結果

学習は 10 エポック付近で収束傾向が見られたため、一つのデータ群につき 10 エポック学習を回した。学習は各データ群につき、RTX-3070 の一般的な PC で約 4 時間かかった。図 1 は、各データ群を学習に使用したモデルと、各テストデータ群の一致率を示したグラフである。各テストデータ群に対して、最も性能が良いモデルの一致率は 51.0%～52.0% であった。先行研究 [4] では、初級者では 44.9%、中級者で 47.3%、上級者で 50.0% の一致率だったことを考えると、単純な比較はできないとはいえ、将棋においても Maia の手法で一致率を十分高められることが分かった。また、グループ 0、グループ 5 のデータを評価に使用した場合、予測モデルの性能が比較的悪いことが分かる。これについては、グループ 0、グループ 5 のレート帯の幅が広いことによる予測が難しいという理由や、棋力の低いプレイヤーには不規則な手が、棋力が高いプレイヤーには探索による手が多いことによる理由が挙げられる。

グラフから、各モデルが学習に使用したデータ群と、最も一致率が高いテストデータ群はある程度一致している、つまり棋力帯ごとの特徴をある程度学習できていることも分かった。例えば、グループ 0 で学習したモデル (青色の折れ線) は、グループ 0 やグループ 1 といった低棋力帯の指し手との一致率が高い代わりに、グループ 4 やグループ 5 といった高棋力帯の指し手との一致率は低い。ここから、グループ 0 で学習したモデルは低棋力帯の指し手のある程度再現できることが分かる。対象が将棋とチェスで異なっているが、本論文で得た結果は Maia の論文で得られた結果と傾向が似ている。Maia の手法に探索を用いた先行研究 [3] では、棋力に応じて探索結果を利用する割合が異なっており、棋力が高いほど探索結果を利用することで一致率がより向上していた。本研究でも棋力によって結果が異なるかを確認するためにグループ 0 を低レート群、グループ

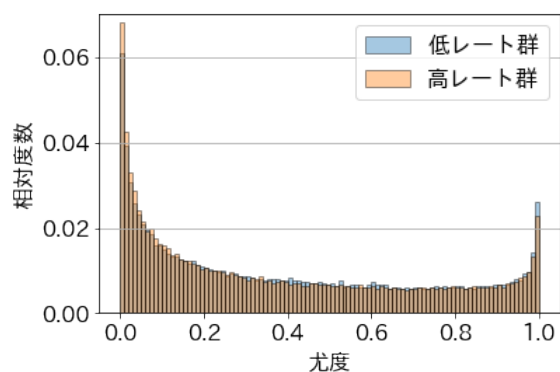


図 2 尤度のヒストグラム

5 を高レート群としてより詳しく分析を行う。

ここまでは先行研究との比較のため一致率についての分析を行っていたが、人間の方策を模倣することを目標と考えるなら、3章で述べたように尤度を高めることが重要であるため、以降は尤度についての分析を行う。図2は、低レート群、高レート群の尤度のヒストグラムである。相乗平均尤度は、低レート群で0.191、高レート群で0.176であった。低レート群と高レート群で分布が非常に似ているが、低レート群では尤度が高い局面の割合が比較的多く、高レート群では尤度が低い局面の割合が比較的多い。両群ともに端に山があり、尤度が高くモデルが人間の着手を予測しやすい局面と、尤度が低くモデルが人間の着手を予測しにくい局面が多い。特に尤度が0~0.01の局面の数が最も多く、将棋の平均合法手数約120手であることを踏まえると、これらの局面はほとんど予測できていないといえる。もしこれらの局面の尤度を高めることができれば、人間の着手を模倣できるモデルに大きく近づけよう。

4.3 尤度が高い局面・低い局面の分類

図2の山になっていて頻度が高い局面について、どのような局面が含まれているかを実際に確認することで、より予測モデルへの理解が深まり、尤度の改善につながると考えたため、考察・分類を行った。

尤度が0.99~1.00と高く、モデルが予測しやすい局面については、十分予測できているため特に問題はない。具体的にどのような局面で尤度が高くなるかを確認したところ、合法手が極めて少ない局面や、一手詰、手筋の一部分といった局面が多く、容易に高い確率を予測できることは自然だと感じた。

尤度が0.00~0.01と低く、モデルが予測しづらい局面について、以下の5つの理由により分類した。

- 主に着手が予測ミス理由となっているもの
 - － モデルが探索しないことによるミス

人間が少し探索を行えば数手先に大きく形勢を損ねてしまうとすぐわかるような手だとしても、探索を

行わないモデルにとってはそれが判断できず、その手が筋の良い手でありさえすれば選択してしまう可能性が高い。このような手があると、探索が必要な手についての予測確率が低くなり、結果として尤度が低くなる。モデルが探索しないことによるミスの局面については、探索が必要かを予測し、必要だと判定した場合に探索を利用した着手確率分布を用いることで尤度を高めることができると考えている。

- － 人間の探索に関するミス

人間が探索を行う際、うっかりや局面の誤認識によってミスをしてしまう場合がある。例えば、大駒の利きのうっかりや持ち駒の誤認識である。そうした人間の悪手は予測確率が低くてもおかしくない。これらの手は人間らしい見落としの再現につながる可能性がある。人間の探索に関するミスについては、モデルが探索を行う際に局面や持ち駒にノイズを入れたり、角の利きに入っても取られないようにしたりする、という方法で尤度を高めることができると考えている。
- － クリックミス

クリックミスについては、ゲームの外で起こることなので、着手予測モデルは正しく推定できず、予測確率が低くなる。この分類については再現する必要性を感じておらず、数も少ない。
- 主に局面が予測ミスの要因となっているもの
 - － 手が広い局面

手が広い局面では、個々の指し手の着手確率も低くなり、尤度が小さくなりやすい。実際に指される手は一つなので、一致率も低くなる。
 - － 敗勢の局面

どうしてもなく敗勢の局面では粘る手を選ぶ人間もいれば、きれいに斬られる手を選ぶ人間もあり、本質的には手が広い局面に含まれる可能性がある。敗勢の局面については、予測勝率が低い局面では着手確率分布を平坦化させる、といった方法で尤度を高められるのではないかと考えている。また、尤度と直接は関係ないが、粘っても勝ち目のない局面で粘らずに、逆転を目指した形作りをするような将棋AI研究も面白いかもしれない。



図 3 モデルが探索をしないことが原因の局面
先手番
予測勝率：0.37
予測上位 5 手：[▲ 4 八同銀 (0.430), ▲ 2 三歩 (0.187), ▲ 6 八銀 (0.187), ▲ 6 一飛 (0.036), ▲ 4 八飛 (0.035)]
実際に指された手：▲ 6 五歩 (0.00123)



図 5 クリックミスが原因の局面
後手番
予測勝率：0.262
予測上位 5 手：[△ 7 九飛成 (0.338), △ 5 八金 (0.186), △ 8 八飛成 (0.101), △ 7 五飛成 (0.081), △ 7 六飛成 (0.070)]
実際に指された手：△ 9 八飛成 (0.00041)



図 4 人間の探索に関するミスが原因の局面
後手番
予測勝率：0.489
予測上位 5 手 [△ 5 三桂 (0.968), △ 4 四玉 (0.020), △ 3 三玉 (0.005), △ 5 三玉 (0.005), △ 5 三銀 (0.000)]
実際に指された手：△ 3 三玉 (0.00532)

分類した局面の中で、典型的なものを取り上げる。図 3 は、予測モデルが探索をしないことを原因に分類した局面である。4 八銀と打たれたところで、5 七の銀取りをどう受けるかという局面である。一見 4 八同銀が自然に見え、予測モデルもこの手を一番に予測している。しかしこの手を読み進めると、4 八同銀、同歩成、同飛車に 5 七角で王手飛車を食らう (局面評価値：-2257, 水匠 5 で 1000 万ノード探索)。先手玉は飛車を渡すと危ない形であるので、王手飛車まで読めるのであれば 4 八同銀は指しづらい。実際に指された手は 6 五歩 (局面評価値：-477, 水匠 5 で 1000 万ノード探索) と水匠の示す最善手で、馬を利かせて銀取りを防ぎつつ、王手飛車もかからない手である。

図 4 は、人間の探索に関するミスを原因に分類した局面



図 6 手が広いことが原因の局面
先手番
予測勝率：0.633
予測上位 5 手：[▲ 3 六歩 (0.221), ▲ 6 五歩 (0.216), ▲ 1 三歩成 (0.136), ▲ 6 八銀 (0.084), ▲ 2 八香 (0.045)]
実際に指された手：▲ 7 七角 (0.00406)

である。5 三金と指されたところだが、これを 5 三同桂と取れば脅威になっている金と成香を排除でき、予測モデルもこの手を 0.97 という高い確率で予測している。実際に指された手は 3 三玉で、おそらく自身の桂馬の利きを見落としている。5 三同桂の局面は後手勝勢 (局面評価値：-2718, 水匠 5 で 1000 万ノード探索) なのに対して、3 三玉は五角 (局面評価値：213, 水匠 5 で 1000 万ノード探索) である。

図 5 は、クリックミスを原因に分類した局面である。6 九銀と飛車に当てられたところで、飛車を逃がす手か 5 八金と取る手が有力である。実際に指された手は 9 八飛成で、おそらく 8 八飛成と指すところをクリックミスで 9 八を選択している。

図 6 は手が広いことを原因に分類した局面である。3 一



図 7 敗勢が原因の局面

後手番

予測勝率: 0.111

予測上位 5 手: [△ 7 七銀不成 (0.579), △ 8 四歩 (0.189), △ 7 七銀成 (0.068), △ 4 六竜 (0.035), △ 6 三角 (0.034)]

実際に指された手: △ 7 五銀 (0.00869)

角と引いたところで、相手の駒の効率を悪くする手、自身の駒の効率を良くする手、中央を厚くする手、駒得を狙う手、白玉を固める手など、手の選択肢の幅が広い。実際に指された手は 7 七角と白玉を固める手で自然ではあるものの、手の広さから尤度が低くなっている。水匠 5 で 1000 万ノードの探索を行ったところ、最善手 7 七角 (評価値 600)~10 番手 5 九金 (評価値 374) と実際に様々な手が有効である。

図 7 は敗勢を原因に分類した局面である。6 三金と駒を取られたところで、後手玉は受けがなく敗勢の局面である。このような局面では、予測モデルは 7 七銀不成, 8 四歩などの手を高い確率で予測しているが、プレイヤーは他のどんな手を指しても結局負けるので、どんな手もありうる。実際に指された手は 7 五銀と銀を捨てる手で、王手をかけるとすればこの手か 7 七銀不成, 9 七銀不成ぐらいしかない。

5. 実験 2: 確信度が低い局面の一致率の向上

5.1 実験 2 の内容

実験 2 では、3 章で触れた確信度を利用して、確信度が低い局面では手が広いために、探索を利用した評価が高い手を選ぶことで一致率を高められるのではないかと、いう予想を確かめる。まず、実験 1 で使用した低レート群、高レート群に対して、 $\max_k p(C_k|x)$ の値がそのまま確信度として使用可能かを確認する。その後、合議アルゴリズムとして複数の予測モデルを使用し、確信度が低い局面について、探索を利用したモデルの評価値を利用して一致率を高められるかを調査する。使用する予測モデルは、実験 1 で学習したモデルの Policy を利用した Policy Max($\arg \max_k p(C_k|x)$), 水匠 (水匠 5), dlshogi (第 31 回世界コンピュータ将棋選手権バージョン) を採用した。水匠は探索ノード数を 10 万ノード、dlshogi は探索ノード数

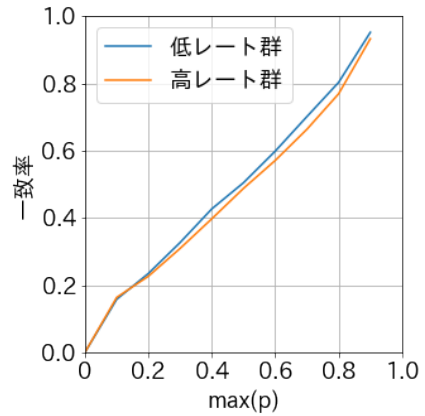


図 8 $\max_k p(C_k|x)$ と一致率の関係

表 1 確信度が低い局面での各予測モデルの予測勝率ごとの一致率 (低レート群)

予測勝率/予測モデル	Policy Max	dlshogi	水匠
低	25.7 %	14.1 %	18.5 %
中	26.1 %	23.3 %	23.0 %
高	27.6 %	27.9 %	27.1 %

表 2 確信度が低い局面での各予測モデルの予測勝率ごとの一致率 (高レート群)

予測勝率/予測モデル	Policy Max	dlshogi	水匠
低	26.1 %	22.2 %	26.0 %
中	26.8 %	30.7 %	30.1 %
高	27.2 %	32.7 %	31.8 %

を 1 万ノードで探索を行った。

局面については、低レート群、高レート群それぞれに対して、実験 1 で使用した予測モデルの確信度が下位 3 分の 1 である、 $\max_k p(C_k|x)$ が 0.4 以下の局面を使用した。これらの局面を実験 1 で使用したモデルの予測勝率で場合分け (低, 中, 高) をして、それぞれ 1 万局面ずつをテストデータとして使用した。

5.2 実験 2 の結果

図 8 は、 $\max_k p(C_k|x)$ と一致率の関係を示した図である。3 章で述べたように、確信度は「自信過剰」な状態、つまり $\max_k p(C_k|x)$ の値が一致率よりも高くなる、ということがしばしば起きる。しかし本研究においては、低レート群においても、高レート群においても $\max_k p(C_k|x)$ の値が一致率と等しくなっているため、 $\max_k p(C_k|x)$ の値をそのままモデルの確信度として解釈可能である。

表 1, 表 2 は、低レート群、高レート群それぞれに対して、確信度が低い局面での各予測モデルの一致率を勝率によって分類したものである。低レート群では Policy Max の一致率を大きく上回る予測モデルは存在しないが、高レート群では形勢が勝勢に近づくにつれて、dlshogi や水匠が Policy Max の一致率を大きく上回った。高レート群で

は、着手を予測する際に良いパラメータで探索することで、より一致率を上げることができるという先行研究の結果とも合致する。ここでは一致率について述べたが、尤度を計算する場合は評価値から確率分布へ変換する方法が必要になる。例えば、各合法手の評価値が分かる場合は、そこからソフトマックス関数と温度 T による温度スケールリング

$$p(C_k|x) = \frac{\exp(u_k/T)}{\sum_{j=1}^K \exp(u_j/T)}$$

によって評価値から確率分布に置き換える方法が挙げられる。

本研究ではさらに、探索を利用する価値が高い局面についてもある程度予測可能ということが明らかになった。この結果については、手が広く見える局面でも、強い人間であれば探索をすることで優れた一手を見つけ出せることが理由だと考えている。その優れた手を Policy Max で予測することが比較的難しい理由は、Policy Max は駒を取ったり当たりになっている駒を逃がしたりする、分かりやすい手を当てることはできるが、数手後にならないと効果が分からないような難しい手を当てることは難しく、互角から勝勢かつ手が広い局面では、そのような状況がよく起きるからではないかと予想している。

予測勝率が低い場合に探索を行っても一致率が上がらない理由は、形勢が悪い場合に人間が相手の間違いを期待する手、相手が読んでいなさそうな手を指すのに対し、探索に基づく最善手の決定ではそのようなことを考慮しないため、指し手が一致しないのではないかと推測している。例えば、AI は負けを先延ばしにするだけの手を指すが、人間はあまり指さないといったことが挙げられる。

6. おわりに

本研究では、実験 1 で Maia の手法は将棋でも有効であることをまず確認した。また、予測モデルの尤度が高い局面や低い局面について、考察・分類を行った。実験 2 では、実験 1 で使用した予測モデルの確信度を利用して、探索によって既存のモデルよりも一致率を向上させられる局面を自動的に判定することに成功した。

将来的には、分類の詳細な定義による自動化、着手確率を出力でき、尤度で比較できるようなモデルの採用、Residual Network 以外で、ConvNeXt などの画像認識でより効果的なネットワークの採用、人間による評価実験を行いたい。

謝辞

本研究では、Mindwalk 株式会社様から「将棋クエスト」の棋譜提供を受けています。深く感謝申し上げます。また、本研究は JSPS 科研費 20K12121 の助成を受けたものです。

参考文献

- [1] 池田心, 楽しませる囲碁・将棋プログラミング, オペレーションズ・リサーチ: 経営の科学, 58(3), 167-173, 2013
- [2] 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
- [3] Jacob, Athul Paul, et al., Modeling strong and human-like gameplay with KLregularized search.” PMLR, 2022
- [4] 杵渕哲彦, 伊藤毅志, 流れを考慮した将棋における人間の指し手との一致率向上手法, 情報処理学会論文誌, 58(9), 1549-1554, 2017
- [5] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... and Hassabis, D., A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science, 362(6419), 1140-1144, 2018
- [6] 杉山卓弥, 小幡拓弥, 斎藤博昭, 保木邦仁, 伊藤毅志, 将棋における合議アルゴリズム—局面評価値に基づいた指し手の選択. 情報処理学会論文誌, 51(11), 2048-2054, 2010
- [7] 岡谷貴之, 機械学習プロフェッショナルシリーズ 深層学習 改訂第 2 版, 講談社, 2022
- [8] Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. On calibration of modern neural networks. In International conference on machine learning, 1321-1330. PMLR, 2017.
- [9] 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

付録：ネットワークの構造

実験 1 で使用した予測モデルの入力形式は、駒の種類ごとの配置と持ち駒を各チャンネルとした、駒の座標を表す多値画像である。入力層はフィルターサイズ 3×3 , フィルター数 192, パディング 1, スライド 1 の畳み込み層とし、活性化関数には ReLU を採用した。中間層は Residual Network を使用し、残差ブロックの各畳み込み層は入力層と同様に、フィルターサイズ 3×3 , フィルター数 192, パディング 1, スライド 1, 活性化関数は ReLU とした。活性化関数の前には Batch Normalization を行った。Residual Network は残差ブロックを 6 ブロック繰り返す形を採用した。方策ネットワークの出力では、移動方向と移動先の座標の組み合わせにより、2187 ラベルの分類を行う多クラス分類問題として捉えた。価値ネットワークの出力では、勝ちを 1, 負けを 0, 引き分けを 0.5 とする二値分類の応用として捉えた*1。最適化手法はモーメントタム SGD を使用し、損失関数には方策ネットワークでは CrossEntropyLoss, 価値関数では BCEWithLogitsLoss を使用した。

*1 この二値分類の実装は山岡の書籍 [9] に従っている。引き分けの数は全体の 0.1 % と少ないため、出力値を勝率とみなすことができると考えている。