Maia2におけるレート帯別モデルの一貫性と棋力差に関する分析

足立 知己[1,a)]　竹内 聖悟[2]

概要：本研究では，人間の棋力差を反映した指し手予測 AI の実現に向けて，Maia2-shogi のレート条件が対局時の棋力および指し手予測に与える影響を分析した．モデル同士の対局実験と将棋倶楽部 24 の評価棋譜を用いた分析を行い，レート帯別に学習した Maia-shogi および Maia2-chess とも比較した．その結果，Maia-shogi では高レートモデルほど低レートモデルに対する勝率が高くなる傾向が見られた一方，Maia2-shogi および Maia2-chess では，本実験条件で明確な棋力差は確認されなかった．また，Maia2-shogi ではレート条件による指し手一致率の変化が小さく，上位 5 手の順位別確率配分の差は序盤で比較的大きかった．さらに，Maia2-shogi は Maia-shogi より Expected Calibration Error（ECE）が小さかった．以上から，本実験条件では Maia2-shogi のレート条件による予測傾向の変化は限定的であり，棋力差や人間の指し手の特徴を再現する方法について，さらなる検証が必要である．

Analysis of Consistency and Playing Strength Across Rating-Level Models in Maia2

Tomoki Adachi[1,a)]　Shogo Takeuchi[2]

Abstract: In this study, with the aim of developing an AI capable of predicting moves that reflect the variation in human playing strength, we analyzed the impact of Maia2-shogi's rating conditions on playing strength during games and move prediction. We conducted experiments involving matches between models and analyzed game records from Shogi Club 24, comparing the results with Maia-shogi and Maia2-chess, which were trained in separate rating bands. The results showed that, while Maia-shogi models with higher ratings tended to have a higher win rate against lower-rated models, no clear difference in playing strength was observed for Maia2-shogi or Maia2-chess under the experimental conditions. Furthermore, in Maia2-shogi, the variation in move agreement rates due to rating conditions was small, and the difference in the probability distribution of the top five moves was relatively large in the opening. Moreover, Maia2-shogi exhibited a smaller Expected Calibration Error (ECE) than Maia-shogi. Based on these findings, under the conditions of this experiment, changes in the prediction trends of Maia2-shogi due to rating conditions were limited; further verification is necessary regarding methods to reproduce differences in playing strength and the characteristics of human moves.

1. はじめに

近年，人間の棋力に応じた実力を再現するゲーム AI の研究が重要となっている．人間の棋力はプレイヤーによって異なり，棋力差は対局時の勝率にも反映される．そのため，人間の棋譜から指し手を学習するだけでなく，プレイヤーの棋力に応じた実力を再現できる AI が求められる．また，各棋力のプレイヤーが実際に指す手の傾向を再現することも，人間らしい AI を実現するうえで重要である．

チェスを対象とした Maia-2 は，プレイヤー自身と対戦相手のレート情報を利用し，単一のモデルで異なるスキルレベルの指し手を予測する手法である．Maia-2 を将棋に適用

[1] 高知工科大学院工学研究科基盤工学専攻
Graduate School of Engineering, Kochi University of Technology
[2] 高知工科大学
Kochi University of Technology
[a)] 305122v@gs.kochi-tech.ac.jp

© 2026 Information Processing Society of Japan

した Maia2-shogi も公開されている．しかし，Maia2-shogi において，レート条件の違いが対局時の棋力差に反映されるか，また，各レート条件が対応する棋力帯の人間の指し手を再現できるかについては，十分に明らかになっていない．

そこで本研究では，Maia2-shogi を対象として，レート条件が対局時の棋力および人間の指し手予測に与える影響を分析し，人間の棋力差を反映した AI として機能するかを検証する．まず，異なるレート条件のモデル同士による対戦実験を行い，勝率から棋力差を評価する．次に，人間の評価棋譜との指し手一致率，一致パターン，指し手確率分布の類似度を分析し，レート条件によって指し手の予測傾向がどの程度変化するかを調べる．さらに，モデルのレート条件と評価棋譜のレート帯の組合せごとに指し手一致率を比較し，対応するレート帯の人間の指し手を再現できるかを検討する．加えて，予測確率と指し手一致率の関係を分析するとともに，レート帯ごとに学習した Maia-shogi およびチェス版の Maia-2 である Maia2-chess との比較を通じて，Maia2-shogi の棋力差と指し手予測の特徴を明らかにする．

2. 関連研究

McIlroy-Young ら [1] は，人間の指し手を予測するチェス AI である Maia を提案した．Maia は，人間の棋譜を用いて学習することで，従来のチェスエンジンよりも高精度に人間の指し手を予測できることを示した．また，プレイヤーの棋力に応じて予測対象を調整することで，特定のスキルレベルにおける人間の指し手を高精度に予測できることを報告している．

Tang ら [2] は，Maia-2 を提案し，従来の Maia で課題となっていたスキルレベル間の予測の不安定さを改善した．Maia-2 では，一つのモデルで複数のレート帯に対応するとともに，自身と対戦相手のレート情報を利用することで，スキルレベルに応じた予測を行う手法を採用している．

Ogawa ら [3] は，将棋において人間らしい指し手を予測する手法を検討し，複数の指し手予測モデルを組み合わせる手法を提案した．同研究では，深層学習による教師あり学習が将棋の指し手予測にも有効であることを示すとともに，棋力の高いプレイヤーの指し手ほど，AlphaZero を基にしたモデルで高精度に予測できることを報告している．また，人間らしさの評価指標として，モデルの予測指し手と人間の指し手が一致する割合を示す指し手一致率と，人間の指し手にモデルが割り当てた確率に基づく尤度を用いている．さらに，複数の予測モデルを状況に応じて選択する Classifier と，各モデルの予測確率を混合する Blend の 2 手法を提案した．実験の結果，Classifier では指し手一致率が 1〜3 ポイント向上した一方，尤度の改善は確認されなかった．これに対し，Blend では指し手一致率が 3〜4 ポイント，尤度が 2〜5 ポイント向上した．

谷合ら [5] は，棋力に応じた将棋の指し手を予測する方策関数を提案した．同研究では，将棋倶楽部 24 の棋譜を用い，自己および相手のレーティングと手数の情報を Adaptive Layer Normalization（AdaLN）によって各層に注入することで，棋力に応じた着手傾向の学習を試みている．実験の結果，人間の着手に対して Top-1 一致率 56.43%，Top-5 一致率 91.40%を達成した．また，棋力条件を用いないベースラインと比較して，全体の予測性能は同等である一方，低レーティング帯ではこれを上回ることを示した．

3. 分析対象

本研究では，モデルに設定するレーティングの値を「レート条件」，評価棋譜を分類するレーティングの範囲を「レート帯」と呼ぶ．

3.1 Maia2-shogi

本研究では，Maia2-shogi を主な分析対象とする．Maia2-shogi は，チェスを対象とした Maia-2 を将棋に適用したモデルである．公開リポジトリの README によると，将棋倶楽部 24 のレーティング 800〜2800 の棋譜から抽出した約 5 億局面を学習に使用している [4]．本研究では，公開リポジトリで配布されている学習済み ONNX モデルを使用する．使用したモデルは，リポジトリの v0.2.1 で公開されているものであり，レート条件 900，1600，2100 を比較する．

Maia2-shogi は，局面とレーティング情報を入力として，指定されたレート条件の人間が指しそうな手を予測する．本研究では，モデル自身のレート条件を 900，1600，2100 に設定した．対戦実験では，対戦相手のレート条件を実際の対戦相手に合わせて設定した．一方，指し手予測実験では，評価対象の棋譜のレート帯に応じて対戦相手のレート条件を設定し，低・中・高レート帯にそれぞれ 900，1600，2100 を用いた．なお，公開リポジトリおよび関連文献 [5] では，ネットワーク構造やレーティング情報の具体的な入力方法について詳細を確認できなかったため，本研究では確認できる範囲の情報を記載する．

3.2 Maia-shogi

Maia-shogi は，文献 [6] の学習プログラム train.py を用いて学習したモデルである．ネットワークには，同プログラムで使用されている PolicyValueNetwork を用いた．このネットワークは，10 個の残差ブロックからなる ResNet を基盤とし，局面から指し手の確率分布を予測する方策出力と，局面における勝敗を予測する価値出力を持つ．本研究では，主に方策出力を用いて，人間の棋譜との指し手の一致を分析する．

学習データには将棋倶楽部 24 の棋譜を使用した．各棋譜

© 2026 Information Processing Society of Japan

の対局者 2 名のレーティングの平均値を基準として，800〜1000，1500〜1700，2000〜2300 の 3 つのレート帯に分類し，各レート帯の棋譜を用いて個別にモデルを学習した．学習データは各レート帯で約 367 万〜480 万局面であり，学習条件はエポック数 10，バッチサイズ 512，学習率 0.001 とした．

3.3 Maia-2

Maia-2 は，チェスを対象とした人間の指し手予測モデルである [2]．本研究では，Maia2-shogi との比較対象として，Maia-2 を対局実験に使用する．対局実験には，国際通信チェス連盟（ICCF）の公式サイトから棋譜を取得した [7]．各棋譜を 15 手目までに制限し，重複を除去した結果，1147 局の棋譜を得た．

3.4 棋譜データ

指し手予測および確率配分の分析には，将棋倶楽部 24 の棋譜を使用した．各棋譜の対局者 2 名のレーティングの平均値を基準として，800〜1000，1500〜1700，2000〜2300 の 3 つのレート帯に分類した．各レート帯について，棋譜を順に読み込み，局面を抽出した．各レート帯で 20000 局面に達した時点で抽出を終了した．抽出対象は特定の手数までに限定せず，棋譜の序盤から終盤までとした．なお，棋譜の抽出において，対局者間のレーティング差による制限は設けていない．

また，評価棋譜の分類に用いる平均レーティングだけでは対局者間の棋力差を十分に表せない可能性があるため，元データについて対局者間のレーティング差を事後的に調査した．その結果，レーティング差が 200 以下の対局は，平均レート帯 800〜1000 で 40676 局（99.74%），1500〜1700 で 37460 局（99.81%），2000〜2300 で 45551 局（93.29%）であった．ただし，これらは元データ全体に対する集計であり，実際に分析に使用した局面のみに限定した集計ではない．

なお，評価棋譜と Maia-shogi の学習に用いた棋譜，および Maia2-shogi の学習データとの重複については検証していない．

4. 実験手法

4.1 対戦実験

レート条件の違いがモデルの棋力に与える影響を調べるため，Maia-shogi，Maia2-shogi および Maia-2 chess を用いて対戦実験を行う．Maia-shogi では，800〜1000 を低レート，1500〜1700 を中レート，2000〜2300 を高レートとし，各レート帯の棋譜を用いて学習した 3 種類のモデルを対戦させる．一方，Maia2-shogi および Maia-2 chess では，同一の学習済みモデルに異なるレート条件を入力して対戦を行う．Maia2 では，自身のレートを `rating_self`，対戦相手のレートを `rating_oppo` として指定する．`rating_self` には各モデルに割り当てたレート条件を設定し，`rating_oppo` には実際の対戦相手に割り当てたレート条件を設定する．Maia2-shogi では，低・中・高レート条件にそれぞれ 900，1600，2100 を設定する．Maia-2 chess では，800，1600，2400 の 3 種類を設定する．

対戦実験では，重複しない序盤の棋譜から用意した 1000 局面を開始局面として使用し，各局面について先手と後手を入れ替えて対局を行う．各モデルの組み合わせにつき合計 2000 局を実施する．

各手番では，モデルが出力する指し手の予測確率を用い，合法手のうち予測確率が最も高い指し手を選択する．合法手が得られない場合は投了として扱う．

対局の終了条件は，投了，千日手，入玉宣言勝ち，詰み，および最大手数への到達とした．千日手については，引き分け，手番側の勝ち，手番側の負けを区別して勝敗を判定した．最大手数は 512 手とし，その時点で対局が終了していない場合は引き分けとして扱った．

4.2 棋譜一致率

AI のレート条件および評価棋譜のレート帯による指し手予測の違いを調べるため，各モデルの予測指し手と評価棋譜の指し手との一致率を算出する．各局面において，モデルが最も高い確率を割り当てた指し手（Top-1）を取得し，評価棋譜の指し手と一致した局面の割合を求める．分析には中レートおよび高レートの評価棋譜を用いる．低レート棋譜では，指し手のばらつきが大きく，棋譜一致率によるモデル間の比較が難しくなる可能性があるため，本分析の対象から除外した．

Maia2-shogi では，‘rating_self‘ に分析対象とする AI のレート条件を設定し，‘rating_oppo‘ には評価棋譜のレート帯に対応する値を設定した．評価棋譜のレート帯が低・中・高の場合，それぞれ 900，1600，2100 を用いた．

さらに，初期局面からの着手数に基づいて局面を 0〜19 手，20〜39 手，40〜79 手，80 手以降の 4 区分に分類する．ここで，着手数は先手・後手それぞれの着手を 1 手として数える．各区分における一致率を比較し，局面の進行に伴う指し手予測の変化を調べる．

4.3 指し手一致パターン

レート条件による指し手予測の違いを調べるため，Maia-shogi および Maia2-shogi を対象に，各レート条件のモデルと評価棋譜の指し手との一致パターンを分析する．低レート，中レート，高レートの 3 条件について，各モデルの Top-1 の指し手が評価棋譜の指し手と一致した場合を 1，一致しなかった場合を 0 として表し，000〜111 の 8 種類のパターンに分類する．

さらに，初期局面からの着手数に基づいて局面を 0〜19

© 2026 Information Processing Society of Japan

手，20〜39 手，40〜79 手，80 手以降の 4 区分に分類し，評価棋譜のレート帯ごとに各パターンの割合を比較する．これにより，AI のレート条件や局面の進行によって，評価棋譜との一致・不一致の傾向がどのように変化するかを調べる．

4.4 上位候補の順位別確率配分の比較

レート条件による上位候補の順位別確率配分の違いを調べるため，各モデルの予測確率上位 5 手を抽出し，それぞれの確率の合計が 1 となるように正規化したうえで，Jensen–Shannon Divergence（JSD）を算出する．本分析では，上位 5 手を予測確率の高い順に並べ，順位ごとの確率を比較する．したがって，本分析は候補手そのものを対応させた指し手分布の比較ではなく，上位候補の順位別確率配分の比較を目的とする．

JSD は次式で定義する．

$$\mathrm{JSD}(P\|Q) = \frac{1}{2}D_{\mathrm{KL}}(P\|M)+\frac{1}{2}D_{\mathrm{KL}}(Q\|M), \quad M = \frac{P+Q}{2}$$

ここで，P および Q は比較対象の上位 5 手の確率を正規化した配列，M は両者の平均，D_{KL} は Kullback–Leibler divergence を表す．対数の底には 2 を用いるため，JSD の値域は 0〜1 であり，値が小さいほど順位ごとの確率配分が類似していることを示す．なお，順位が同じ候補手が同一の指し手であるとは限らないため，JSD の値を指し手分布そのものの類似度として解釈することはできない．

まず，評価棋譜のレート帯ごとに，Maia2-shogi のレート条件 900，1600，2100 の組み合わせについて JSD の平均値を算出する．また，900 と 2100 の比較について，JSD が 0.01，0.02，0.05 以上となる局面の割合を求める．さらに，初期局面からの着手数に基づいて局面を 0〜19 手，20〜39 手，40〜79 手，80 手以降の 4 区分に分類し，各区分における JSD が 0.01 以上となる局面の割合を比較する．

加えて，Maia2-shogi のレート条件 900 と 2100 の JSD が 0.01 以下および 0.001 以下の局面を抽出し，3 つのレート条件の Top-1 の指し手がすべて評価棋譜の指し手と一致する局面の割合（111 率）を算出する．これにより，上位候補の順位別確率配分の違いと，3 つのレート条件における指し手一致との関係を調べる．

さらに，Maia2-shogi と Maia-shogi の予測傾向を比較するため，3 つのレート条件のモデル間で Top-1 の指し手がすべて一致する局面を抽出する．抽出した局面について，同じレート帯に対応するモデル間で JSD を算出する．また，Maia-shogi の 3 つのレート帯のモデル間についても，同様に 3 モデルの Top-1 の指し手がすべて一致する局面を対象として JSD を算出し，モデル間の順位別確率配分の違いを比較する．



図 1 モデルごとの勝率

4.5 予測確率と指し手一致率の関係

モデルの予測確率と評価棋譜の指し手との一致率の対応を調べるため，Top-1 予測確率に基づく分析を行う．全局面を Top-1 予測確率に応じて 0.1 刻みの確率帯に分類し，各確率帯における平均予測確率と評価棋譜との指し手一致率を算出する．確率帯は下端を含み，上端を含まない区間とし，最後の区間は 1.0 を含むものとする．

さらに，予測確率と指し手一致率の差を定量的に評価するため，Expected Calibration Error（ECE）を算出する．ECE は，各確率帯に含まれる局面数の全局面数に対する割合を重みとして，平均予測確率と指し手一致率の絶対差を加重平均した値である．ECE が小さいほど，予測確率と評価棋譜との指し手一致率の差が小さいことを示す．

5. 実験結果

5.1 対戦実験

図 1 に，Maia-shogi，Maia2-shogi および Maia2-chess の対戦結果を示す．Maia-shogi では，高レートモデルの低レートモデルに対する勝率は 62.4%，中レートモデルに対しては 58.5%であり，中レートモデルの低レートモデルに対する勝率は 54.7%であった．このように，Maia-shogi ではレート帯が高いモデルほど，低いレート帯のモデルに対する勝率が高くなる傾向が見られた．

一方，Maia2-shogi では，高レート条件のモデルの中レート条件および低レート条件のモデルに対する勝率は，それぞれ 51.6%，53.3%であった．また，中レート条件のモデルの低レート条件のモデルに対する勝率は 50.3%であった．Maia2-chess でも，高レート条件のモデルの低レート条件および中レート条件のモデルに対する勝率は，それぞれ 50.9%，51.9%であり，中レート条件のモデルの低レート条件のモデルに対する勝率は 49.1%であった．Maia2-shogi と Maia2-chess では，いずれの組み合わせでも勝率は 50%前後であり，レート条件の違いに応じた明確な勝率差は見られなかった．

5.2 棋譜一致率

対局実験において Maia2-shogi ではレート条件による棋力差がほとんど確認されなかった要因を調べるため，レー

© 2026 Information Processing Society of Japan

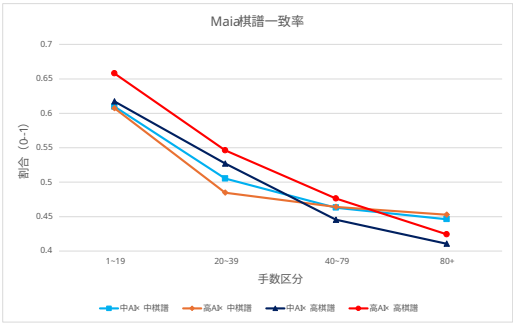


図 2　Maia-shogi のレート条件別棋譜一致率

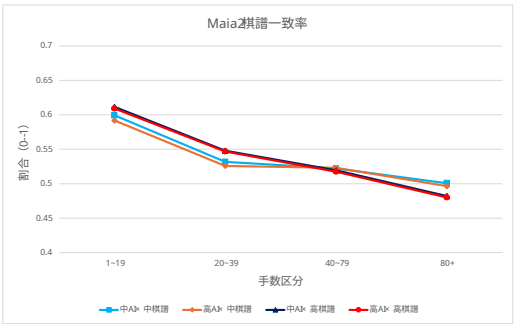


図 3　Maia2-shogi のレート条件別棋譜一致率

ト条件ごとの Top-1 棋譜一致率を比較した．低レート棋譜では一致率の変動が大きかったため，本分析では中レートおよび高レートの棋譜を対象とした．

図 2 に Maia-shogi，図 3 に Maia2-shogi の棋譜一致率を示す．Maia-shogi では，高レート棋譜に対して高レートモデルの一致率がすべての手数区分で高く，特に 40〜79 手では，中レートモデルの 44.5%に対して高レートモデルは 47.6%であった．一方，中レート棋譜では，高レートモデルの一致率が一貫して高くなる傾向は見られなかった．

Maia2-shogi では，中レートおよび高レートの棋譜のいずれに対しても，レート条件の変更による一致率の差は小さかった．例えば，高レート棋譜の 40〜79 手では，中レート条件が 51.9%，高レート条件が 51.7%であり，高レート条件による改善は確認できなかった．

以上から，Maia-shogi では高レート棋譜に対して高レートモデルの一致率が高くなる傾向が見られた一方，Maia2-shogi ではレート条件と棋譜レートの対応による明確な差は確認できなかった．ただし，本分析では低レート棋譜を対象としていないため，低レート条件と低レート棋譜の対応関係については評価できていない．

5.3　指し手一致パターン

棋譜一致率だけではモデル間の違いを把握しにくいため，レート変更によって評価棋譜との一致・不一致の組合せがどの程度変化するかを調査した．その結果，Maia-shogi では一部のモデルのみ一致となる局面が一定数見られた一方，Maia2-shogi ではその割合が小さかった．

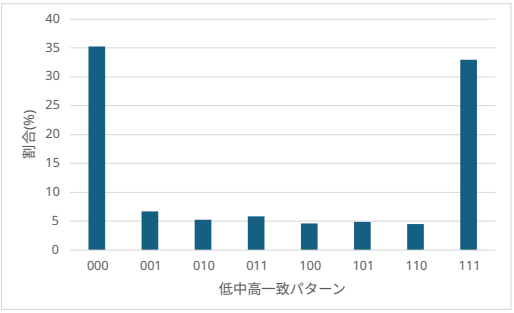


図 4　指し手一致パターン分布（Maia-shogi）

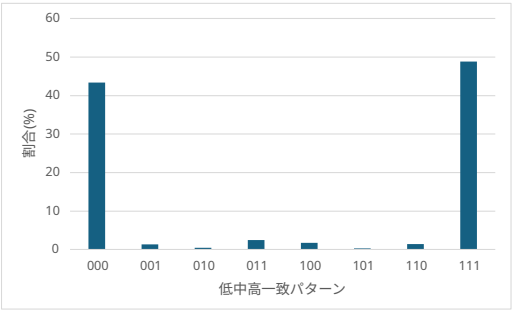


図 5　指し手一致パターン分布（Maia2-shogi）

低・中・高レートの評価棋譜を対象として，Maia-shogi および Maia2-shogi の低・中・高レートモデルの指し手を比較した．評価棋譜のレート帯による影響は小さかったため，高レートの評価棋譜に対する結果を図 4 および図 5 に示す．評価棋譜との一致を 1，不一致を 0 として表し，000〜111 の 8 種類の一致パターンに分類した．さらに，これらを「全モデル一致」(111)，「全モデル不一致」(000)，「一部のモデルのみ一致」(001，010，011，100，101，110) の 3 種類に分類した．

分析の結果，Maia-shogi，Maia2-shogi ともに「全モデル一致」および「全モデル不一致」の割合が高かった．一方，「一部のモデルのみ一致」に該当する 6 パターンの合計割合は，Maia-shogi では約 35%であったのに対し，Maia2-shogi では約 4%にとどまった．この傾向は他の評価棋譜のレート帯でも同様であった．

5.4　上位候補の順位別確率配分の比較

Maia2-shogi のレート条件間における上位候補の順位別確率配分の違いを調べるため，レート条件 900，1600，2100 の組み合わせごとに JSD を算出した．評価棋譜のレート帯ごとの平均 JSD を表 1 に示す．

表 1 より，レート 900 と 1600，および 1600 と 2100 のモデル間の平均 JSD は約 0.002〜0.004 であった．一方，レート 900 と 2100 のモデル間では約 0.0073〜0.0075 となり，隣接するレート条件間より大きかった．この傾向は，評価棋譜のレート帯によらず共通していた．

次に，レート 900 と 2100 のモデル間について，JSD が各閾値以上となる局面の割合を調べた結果，JSD が 0.01 以

© 2026 Information Processing Society of Japan

表 1　評価棋譜のレート帯別に見た Maia2-shogi の平均 JSD

棋譜レート	900–1600	1600–2100	900–2100
800–1000	0.003934	0.001965	0.007342
1500–1700	0.004050	0.002033	0.007506
2000–2300	0.003857	0.002073	0.007265

表 2　Top-1 一致局面におけるモデル間の平均 JSD

対象局面の条件	比較モデル	平均 JSD
Maia2-shogi が 111	Maia900–DL800–1000	0.040727
Maia2-shogi が 111	Maia1600–DL1500–1700	0.040902
Maia2-shogi が 111	Maia2100–DL2000–2300	0.041968
Maia-shogi が 111	Maia900–DL800–1000	0.039080
Maia-shogi が 111	Maia1600–DL1500–1700	0.038928
Maia-shogi が 111	Maia2100–DL2000–2300	0.043103
Maia-shogi が 111	DL800–1000–DL1500–1700	0.036053
Maia-shogi が 111	DL1500–1700–DL2000–2300	0.036258
Maia-shogi が 111	DL800–1000–DL2000–2300	0.039570

上となる局面は約 18〜19%，0.02 以上では約 9%，0.05 以上では約 2%であった．また，JSD が 0.01 以上となる局面の割合は，0〜19 手では約 56〜59%，20〜39 手では約 34〜37%であったのに対し，40〜79 手では約 6%，80 手以降では約 1〜2%であった．このことから，上位候補の順位別確率配分におけるレート条件間の差は序盤に多く見られ，手数が進むにつれて小さくなる傾向が確認された．

さらに，Maia2-shogi の 3 つのレート条件間で Top-1 の指し手がすべて一致する局面を対象に，Maia2-shogi と Maia-shogi の確率配分を比較した．また，Maia-shogi の 3 つのレート帯のモデル間についても，3 モデルの Top-1 の指し手がすべて一致する局面を対象に JSD を算出した．これらの結果を表 2 に示す．

Maia2-shogi の 3 条件で Top-1 の指し手が一致する局面では，同レート帯に対応する Maia2-shogi と Maia-shogi のモデル間の平均 JSD は約 0.041 であった．一方，Maia-shogi の 3 モデルで Top-1 の指し手が一致する局面では，Maia2-shogi と Maia-shogi のモデル間で約 0.039〜0.043，Maia-shogi のモデル間で約 0.036〜0.040 であった．このように，Top-1 の指し手が一致する局面に限定しても，モデル間の順位別確率配分には差が見られた．ただし，本分析では上位 5 手を確率順に並べて比較しているため，順位が同じ候補手が同一の指し手であるとは限らない．

5.5　予測確率と指し手一致率の関係

モデルの予測確率と評価棋譜との指し手一致率の対応を調べるため，Top-1 予測確率に基づく分析を行った．全局面を Top-1 予測確率に応じて 0.1 刻みの確率帯に分類し，各確率帯における平均予測確率と指し手一致率を算出した．また，両者の差を定量的に評価するため，Expected Calibration Error（ECE）を算出した．ECE は，各確率帯の局面数に応じて，平均予測確率と指し手一致率の絶対差を加重平均した値であり，小さいほど両者の対応がよいことを示す．

表 3　レート条件を対応させたモデルの ECE

教師棋譜のレート帯	モデル	ECE（ポイント）
800–1000	Maia-shogi	7.98
	Maia2-shogi	2.41
1500–1700	Maia-shogi	7.74
	Maia2-shogi	1.61
2000–2300	Maia-shogi	7.52
	Maia2-shogi	1.67

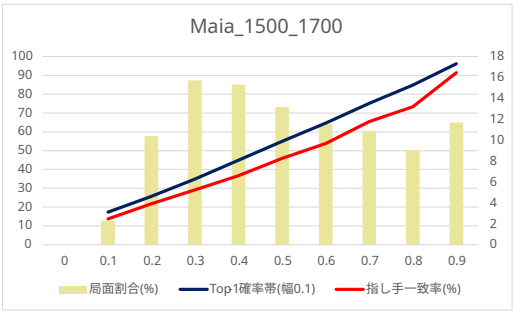


図 6　予測確率と指し手一致率の関係（Maia-shogi，教師棋譜レート 1500–1700）

図 6 および図 7 に，教師棋譜のレート帯が 1500–1700 の場合における，Maia-shogi および Maia2-shogi の予測確率と指し手一致率の関係を示す．また，表 3 に，各教師棋譜レート帯における対応するモデルの ECE を示す．Maia-shogi の ECE は 7.52〜7.98 ポイントであったのに対し，Maia2-shogi の ECE は 1.61〜2.41 ポイントであり，すべてのレート帯で Maia2-shogi の方が小さかった．この結果から，今回の評価条件では，Maia2-shogi の方が Top-1 予測確率と評価棋譜との指し手一致率がよく対応していることが示された．

ただし，ECE は予測確率と指し手一致率の対応を評価する指標であり，レート条件に応じた指し手の再現性や棋力を直接評価するものではない．したがって，Maia2-shogi の ECE が小さいことのみから，レート帯ごとの棋力や指し手の特徴をより適切に再現しているとは結論づけられない．

6.　考察

6.1　レート条件と対局時の棋力の関係

Maia-shogi では高レートモデルほど低レートモデルに対する勝率が高くなった一方，Maia2-shogi および Maia2-chess では，本実験条件では明確な棋力差は確認されなかった．この違いには，レート帯ごとに個別のモデルを学習する方式と，単一のモデルにレート条件を入力する方式の違いが関係している可能性がある．ただし，学習データやモデル構造も異なるため，原因の特定には追加検証が必要で

© 2026 Information Processing Society of Japan

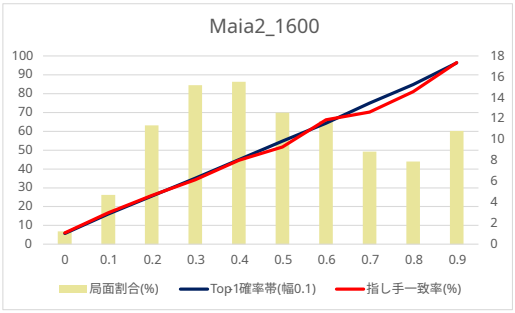


図 7 予測確率と指し手一致率の関係（Maia2-shogi，教師棋譜レート 1500–1700）

ある．

また，Maia-2 は人間の指し手予測を目的とするため，レート条件によって予測傾向が変化することと，モデル同士の対局で勝率差が生じることは必ずしも一致しない．そのため，対局結果だけでなく，評価棋譜との指し手一致率や予測確率分布も併せて評価する必要がある．

なお，本研究の対局実験では，各局面で予測確率が最大の指し手を選択した．確率に応じて指し手を選択する方法など，異なる選択方法を用いた場合に結果が変化する可能性があるため，今回の結果のみから Maia2-shogi 自体が人間の棋力差を再現できないと結論づけることはできない．

6.2 レート条件と指し手予測の関係

Maia2-shogi では，レート条件による指し手一致率の変化が小さく，一部のモデルのみが評価棋譜と一致する局面も少なかった．JSD による比較でも，隣接するレート条件間の確率配分の差は小さかった．これらの結果から，今回の評価条件では，レート条件による予測傾向の変化は限定的であったと考えられる．一方，レート条件間の差が大きい場合や序盤では，確率配分に比較的大きな差が見られた．

ただし，レート条件によって予測が変化することと，対応するレート帯の指し手をより正確に予測できることは異なる．後者を判断するには，評価棋譜のレート帯ごとに各レート条件の一致率を比較する必要がある．また，本研究の JSD は上位 5 手を確率順に並べて比較しており，候補手そのものを照合していないため，確率配分の差が具体的な指し手の違いをどの程度反映するかは明らかでない．

6.3 予測確率と指し手一致率の関係

Maia2-shogi は Maia-shogi より ECE が小さく，Top-1 予測確率と評価棋譜との指し手一致率の差が小さかった．これは，今回の評価条件では，予測確率と実際の指し手一致率が比較的よく対応していたことを示す．ただし，この対応は棋力差やレート帯に応じた指し手の再現性を直接示すものではない．したがって，予測確率と一致率の対応，指し手の再現性，対局時の棋力は区別して評価する必要がある．

6.4 本研究の限界と今後の課題

本研究では，各局面で予測確率が最大の指し手を選択する対局実験を行ったが，Maia2-shogi のレート条件間で明確な棋力差は確認されなかった．ただし，この結果は本実験の指し手選択方法に基づくものであり，異なる選択方法を用いた場合については検証していない．今後は，レート条件と評価棋譜のレート帯の対応関係をより詳しく検証するとともに，確率に応じた指し手選択なども比較し，確率配分の違いが実際の指し手選択や対局結果にどう関係するかを分析する必要がある．また，評価棋譜と学習データの重複を確認できていない点も限界であり，評価データの独立性を確認したうえで検証することが望ましい．

7. まとめと今後の展望

本研究では，人間の棋力差を反映した指し手予測 AI の実現に向けて，Maia2-shogi のレート条件が対局時の棋力および指し手予測に与える影響を分析した．異なるレート条件のモデル同士による対局実験に加え，将棋倶楽部 24 の評価棋譜を用いて，指し手一致率，一致パターン，上位 5 手の順位別確率配分の類似度を比較した．また，レート帯ごとに学習した Maia-shogi およびチェス版の Maia2 である Maia2-chess との比較を行った．

実験の結果，Maia-shogi では高レートモデルほど低レートモデルに対する勝率が高くなる傾向が見られた．一方，Maia2-shogi および Maia2-chess では，勝率が 50%前後にとどまり，本実験条件では明確な棋力差は確認されなかった．また，Maia2-shogi ではレート条件による指し手一致率の変化が小さく，一部のモデルのみが評価棋譜と一致する局面の割合も約 4%にとどまった．上位 5 手の順位別確率配分を比較した結果，レート条件間の差は序盤に比較的大きく，手数の進行に伴って小さくなる傾向が見られた．さらに，Maia2-shogi は Maia-shogi より ECE が小さく，予測確率と評価棋譜との指し手一致率がよく対応していた．

以上から，本実験条件では，Maia2-shogi のレート条件による予測傾向の変化は限定的であり，対局時の棋力差も明確には確認されなかった．ただし，対局実験では各局面で予測確率が最大の指し手を選択しているため，今回の結果のみから，Maia2-shogi 自体が人間の棋力差を再現できないと結論づけることはできない．

今後は，Maia-shogi と Maia2-shogi の予測確率を組み合わせる Blend を検討する．Maia-shogi のレート条件に応じた勝率の傾向と，Maia2-shogi の予測確率と指し手一致率の対応のよさを活用し，棋譜との指し手一致率をなるべく維持しながら，レート条件に応じた棋力差を反映できるモデルの実現を目指す．具体的には，両モデルの予測確率を混合する割合を変化させ，指し手一致率と対局時の勝率を比較することで，指し手の再現性と棋力の両立が可能かを検証する．また，チェスでは Maia-3 が公開されている

© 2026 Information Processing Society of Japan

ため，今後は Maia-3 を比較対象に加え，Maia2-chess との比較を通じて，レート条件による棋力差や指し手予測の特徴を分析する．

参考文献

[1] R. McIlroy-Young, S. Sen, J. Kleinberg, and A. Anderson, “Aligning Superhuman AI with Human Behavior: Chess as a Model System,” Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2020), pp. 1677–1687, 2020.

[2] Z. Tang, D. Jiao, R. McIlroy-Young, J. Kleinberg, S. Sen, and A. Anderson, “Maia-2: A Unified Model for Human-AI Alignment in Chess,” Advances in Neural Information Processing Systems (NeurIPS 2024), Vol. 37, 2024.

[3] T. Ogawa, C.-H. Hsueh, and K. Ikeda, “Improving the Human-Likeness of Game AI’s Moves by Combining Multiple Prediction Models,” Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023), Vol. 3, pp. 931–939, 2023.

[4] nyoki-mtl, “Maia-2 Shogi,” GitHub repository, `https://github.com/nyoki-mtl/maia2-shogi`（参照 2026-10-19）.

[5] 谷合廣紀，門本淳一郎，入江英嗣，「条件付き正規化を用いた棋力に応じた将棋方策関数の学習」，情報処理学会研究報告ゲーム情報学（GI），Vol. 2026-GI-57，No. 2，pp. 1–6，2026.

[6] 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

[7] International Correspondence Chess Federation (ICCF), Official Website, `https://www.iccf.com/`（参照 2026-10-19）.

© 2026 Information Processing Society of Japan