修士論文

限られた枚数の棋譜を活用した人間らしい価値関数と方策関数の提案

小川竜欣

主任研究指導教員　　池田心

北陸先端科学技術大学院大学　金沢大学
融合科学共同専攻
（融合科学）

令和 5 年 3 月

限られた枚数の棋譜を活用した人間らしい価値関数と方策関数の提案
(Proposals for human-like value functions and policy functions trained using a limited number of game records)

北陸先端科学技術大学院大学　学生番号 2150004
氏名　小川竜欣
主任研究指導教員氏名　池田心

ゲームを対象とした情報処理に関する研究はゲーム情報学と呼ばれている．ゲーム情報学では長らく，「人間のトッププレイヤに勝つ」という目標が掲げられてきたが，将棋や囲碁，チェスといった完全情報ゲームではこの目標は既に達成された．しかし，人間を楽しませるという点から見ると，残された課題は多い．

人間らしい将棋 AI を実現するにあたって，将棋 AI を構成する要素について着目すると，「価値関数（局面から勝率を予測する）」と「方策関数（局面から着手確率を予測する）」，「探索手法（先読みを行い，価値関数・方策関数の評価を精緻化する）」という構成例が挙げられる．本論文では，探索手法の構成要素となっている価値関数と方策関数の人間らしさを向上させる手法について提案し，実際に人間らしさを向上できているかについて評価を行った．

人間らしい価値関数について述べる．最近ではプロ棋士の対局で局面の評価値が示されることも多いが，人間プレイヤの実感または実際と乖離した評価値が示されることもある．本論文では，局面から勝率を予測する教師あり学習を行う際に，棋力情報も入力に含めることで，より人間らしい局面評価を目指した．また，推定した勝率が実際の勝率に近いかを確かめるため，指し継ぎによる評価を行った．指し継ぎには，形勢判断に人間的な項目を採用している技巧 2 を用いた．探索の深さを制限した弱い将棋 AI による指し継ぎの勝敗は，我々のモデルの予測勝率のほうが，強い将棋 AI の予測勝率よりも近かった．例えば，指し継ぎ結果が先手勝率 0.35 の局面では，強い将棋 AI が勝率 0.89 と予測するところを提案したモデルは勝率 0.23 と予測した．また，同一局面で入力する棋力を変えた場合に，予測勝率が大きく異なる局面をサンプリングして，局面の解釈を行った．その結果，このサンプリング方法で抽出したそれぞれの局面は，たしかに逆転が起こりやすい局面であろうことを確認できた．

人間らしい方策関数について述べる．強いゲーム AI が調整なしで人間と対局すると棋力が高すぎるため，ランダムな行動をとらせたり探索を浅くしたりといった単純な方法で弱体化させることがある．これらのゲーム AI の行動は，ときに人間にとって奇妙であったり，理解しがたいものであったりする．これは，単に対局して人間に勝利したり，互角の勝負をしたりすることだけが目的であれば問題ないが，人間を楽しませることを目的とする場合に問題になる．なぜなら，人間プレイヤはゲーム AI の手が不自然である，もしくは理解できないと感じると，対局を楽しむのは難くなるためである．

この問題を解決するため，チェスや囲碁で人間の着手予測に有効な手法として知られている深層教師あり学習モデル[1]と，強いゲーム AI を作る手法として知られている AlphaZero[2]系の強化学習モデルについて，各モデルが将棋の着手予測についても有効か調査を行った．その結果，1 モデルにつき約 13 万棋譜を使用した深層教師あり学習は人間の手を 50%程度予測でき，将棋においても有効な手法であることを示した．強化学習に基づく AlphaZero 系の将棋 AI である DLshogi は，棋力が高いプレイヤの手をより正確に予測できることを示した．これらの 2 つのモデルはそれぞれ強みがあるため組み合わせることが有望だと考えたが，異なる視点からも分析を行うことで，モデルについて理解をより深められると考えた．

そこで，深層教師あり学習モデルについて，尤度（モデルによって予測される人間の手の確率）に着目して分析を行った．そこから，低棋力プレイヤのデータでは予測確率が 0.01 以下の人間の手は 5%ほど存在するなど，モデルが予測できていない人間の手が少なからず存在することを示した．このようなことが起きる理由を調べるため，モデルが予測しづらい局面について考察・分類を行った．分類については，「モデルが探索しないことによるミス」，「人間の探索に関するミス」，「操作ミス」，「様々な手が有望な局面」，「敗勢の局面」という 5 つに分け，各分類ごとに尤度を高める手法について提案を行った．

また，人間らしい方策関数については，教師データの数が限られている場合に，複数の方策を組み合わせることで，一致率・尤度の向上を目指す 2 つの手法を提案した．一つは，Classifier モデルという，異なる状況に応じて適切な方策関数を選択する「分類器」を用いるものであり，もう一つは，Blend モデルという，複

数の方策関数の確率を「混合」するものである．実験の結果，Classifier モデルでは一致率については低棋力プレイヤのデータでは 1.3 ポイント，高棋力プレイヤのデータでは 2.0 ポイント向上したが，尤度については向上しないことが分かった. Blend モデルでは, 一致率については低棋力プレイヤのデータでは 2.9 ポイント，高棋力プレイヤのデータでは 3.7 ポイント向上した．また，尤度についても低棋力プレイヤデータでは 0.200 から 0.224，高棋力プレイヤのデータでは 0.201 から 0.224 に改善した．このように，Classifier モデルも Blend モデルも一致率を向上させることに成功したが，Blend モデルのほうが一致率の向上幅が大きいうえ，尤度の向上にも成功したため，より優れた手法と言える．

本研究の結果をまとめると，人間らしい価値関数については，逆転が起こりやすい局面について自動的に抽出することに成功した．このモデルを，逆転のしやすさについて自身で判断することが難しいプレイヤに利用してもらうことで，上達を支援できるかもしれない．人間らしい方策関数については，強みが異なる深層教師あり学習モデルと強化学習モデルを組み合わせることで，一致率・尤度を改善することに成功した．このモデルをアマチュアの対局の観戦に利用することで，実際に指されやすい手について観戦者が把握し，臨場感を味わうことができるかもしれない．

参考文献（最大 5 件）

[1] McIlroy-Young, R., Sen, S., Kleinberg, J. and Anderson, A.: Aligning super-human AI with human behavior: Chess as a model system, in Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1677–1687 (2020).

[2] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K. and Hassabis, D.: A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, Science, Vol. 362, No. 6419, pp. 1140–1144 (2018).

発表論文・口頭発表

[1] 小川竜欣, 池田心. 対局状況をより正確に表現するための盤面評価値, 第 26 回ゲームプログラミングワークショップ (GPW), pp.28‐33, (2021).

[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, 第 27 回ゲームプログラミングワークショップ (GPW), pp.180‐186, (2022).

[3] Ogawa, T., Hsueh, C.-H., Ikeda, K.: Improving the Human-Likeness of Game AI 's Moves by Combining Multiple Prediction Models, 15th International Conference on Agents and Artificial Intelligence (ICAART), Paper #276, (2023)

目 次

図 目 次

表 目 次

第1章　はじめに

ゲームを対象とした情報処理に関する研究はゲーム情報学と呼ばれている．ゲームは,「ルールが明確であり，実装を行いやすい」,「提案された手法がどれほど有効かを，勝敗や得点によって明確に評価できる」,「プレイデータが収集しやすく，強い人間プレイヤが存在することも多い」といった特徴を持っており，人工知能の研究対象として良い題材である．また，推論，記憶，学習など，さまざまな人間の知的行動が含まれており，認知科学の研究対象としても優れている．このように，ゲーム情報学は人工知能（AI）研究，認知科学研究において重要な地位を占めてきた [1].

ゲーム情報学が特に注目するのは，AI・人間プレイヤの行動部分である．ゲームをプレイするためには何らかの知的行動が必要であり，人工知能の観点からは「知的行動をどのようにして計算機上で実現するか」，認知科学の観点からは「知的行動をどのようにして人間が実現し，プレイするために何をどのように学習しているか」について着目しているといえる．

ゲーム AI が対戦によって人間プレイヤを楽しませるためには，ある程度の強さがまず求められる．そこでゲーム情報学においては，長らく強いゲーム AI についての研究が主流であり,「人間のトッププレイヤに勝つ」という目標が掲げられてきた．この目標については，チェスでは Deep Blue[2] が，囲碁では AlphaGo[3] が人間のトッププレイヤに勝利し，ある程度達成されたといえる．しかし，チェスや将棋では，従来の強いゲーム AI やそれを弱体化したものは人間と大きく異なる行動をすることが指摘されている [4][5]．これは，人間と対局して楽しませたり，人間にゲームを教えたりするといった，人間とゲーム AI の相互作用を考えたときに問題となる．そこで，一人称シューティングゲームで人間よりも人間らしいと評価されたゲーム AI[6] といった，人間らしいゲーム AI に関する研究も行われている．

将棋は，チェスよりも探索空間が広く，ゲーム AI を強くする難易度が高いと言われてきたが，2017 年，Ponanza が佐藤天彦名人（当時）に 2 勝 0 敗で勝利した．このように，人間と互角以上の対局を行うには十分な強さの将棋 AI が実現している．そこで，人間の棋風の再現を目指した研究 [7] や，人間の手を既存手法よりも正確に予測することを目指した研究 [8] も行われている．しかし，人間よりも人間らしく感じる将棋 AI を実現するにはまだ課題が多く残っている．

ここからは人間らしい将棋 AI を実現するにあたって，将棋 AI を構成する要素に着目する．将棋 AI の構成要素の一例として,「価値関数（局面から勝率を予測する）」,「方策関数（局面から着手確率を予測する）」,「探索手法（先読みを行い，価

値関数・方策関数の評価を精緻化する)」という構成例が挙げられる．我々は，最終的に指し手を決定する探索手法の構成要素である，価値関数と方策関数について人間らしさを向上させる必要があると考えた．そこで，本論文では価値関数と方策関数を研究対象とする．

人間らしい価値関数について，現状使用されている価値関数，もしくは局面評価関数の値は，一般的な人間プレイヤの実感や，人間プレイヤが実際に指した場合の結果と大きく異なる場合がある．これは，同じ局面でもプレイヤの棋力によって勝率・評価が異なるが，予測している将棋 AI にとっての勝率・評価値を一律に出力していることが大きな原因の一つだと考えている．そこで，本研究では，棋力情報を数値として入力に含めた深層教師あり学習によって勝率予測を行うことで，より現実に近い対局状況の表現を目指す．また，初中級者へ勝率をより上げられるアドバイスをするといった応用のため，逆転が起こりやすい局面，起こりにくい局面を自動的に抽出できるかを確かめる．

人間らしい方策関数について，前述のとおり，強いゲーム AI の方策は人間のものと大きく異なっていることが知られており，人間とゲーム AI の相互作用を考えたときに問題になる．本研究では，まず，チェスにおいて着手予測に最も有効な手法の一つとして知られている，深層教師あり学習が将棋でも有効かを検証する．さらに，そのモデルでは予測しづらい局面について分類・考察を行う．このようなことを行う目的は，各分類について人間らしさの評価を向上させる手法を考案するためである．本論文では特に，学習モデルが探索を行わないために起こる予測ミスに注目し，人間の棋譜から教師あり学習したゲーム AI の方策と，棋譜を使用せず強化学習したゲーム AI の方策といった，長所と短所が異なるゲーム AI の方策を組み合わせる 2 つの手法について提案する．

本論文の構成は以下の通りである．第 2 章では研究対象とする将棋のルールや性質，考え方について説明する．第 3 章では関連研究の紹介を行う．第 4 章では人間らしい価値関数についての実験とその結果について述べる．第 5 章では人間らしい方策関数についての実験とその結果について述べる．第 6 章では本研究のまとめについて述べる．

第2章　将棋

本章では将棋の基本的なルールについて述べた後，持ち時間の制度や，対局の段階と戦略について，ウィキペディア [9][10] を参考にして紹介する．

2.1　将棋の基本ルール

将棋は二人でプレイするボードゲームの一種で，一般的に「将棋」という場合には本将棋のことを指す．対局において先に駒を動かし始める側のプレイヤを先手，そうでない側の対局者を後手と呼ぶ．将棋では一局を通じて，先手と後手が互いに自分の駒を 1 つ動かすか，持ち駒（相手から取って自分のものとなった駒）を 1 つ盤上に置くことを繰り返す．玉将と金将以外の自分の駒を動かして，敵陣に入る，敵陣の中で動く，敵陣から出る場合は「成る」ことを選択できる．金将以外の小駒（銀将・桂馬・香車・歩兵）は成ると金将の動きになり，大駒（飛車・角行）は動けなかった方向に 1 マスずつ動けるようになる．成った駒が相手に取られた場合は成る前の状態に戻る．プレイヤは基本的に自身の駒で相手の玉将という駒を捕獲することを目指す．しかし，玉が敵陣に入った場合（入玉）は捕獲することが難しいため，入玉したときは駒を点数化して決着をつけることがある．点数化については，玉将は 0 点，大駒を 1 枚 5 点，小駒を 1 枚 1 点として点数を合計する．

将棋における勝敗の決まり方は以下のようなものがある．

- 片方のプレイヤが以下の状態になった場合には，そのプレイヤは負けとなり，もう一方のプレイヤの勝ちになる．
  - 詰み（自分の玉が次に取られる状況だが，それを避ける手がない.）
  - 投了（勝利することが難しいと判断して負けを認めた.）
  - 反則行為（以下の反則を行ったことを指摘された.）
    - ルール違反（基本ルールに反する動作を行った.）
    - 禁じ手（ルールで禁止された手を指した.）
    - 連続王手の千日手（相手玉への王手の連続によって同一局面が 4 回現れた．図 2.1 は連続王手の千日手の局面例である.）

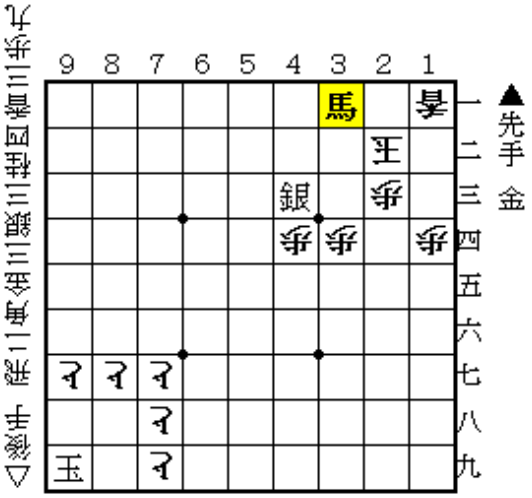


図 2.1: 連続王手の千日手の局面例．この局面では先手は自玉の詰みを防ぐ手がなく，勝利するためには後手玉を詰ませる必要がある．後手は 3 一玉と馬を取ると 3 二金で詰み，1 二玉は 1 三金で詰むが，3 三玉と上がる手がある．先手の有効そうな王手は 4 二馬しかないが，その局面で後手に 2 二玉と指されると先手は 3 一馬とせざるを得ず，その局面は初めの局面と同じで，この手順を繰り返すと連続王手の千日手になり，先手の負けになる．

  - 相入玉での点数不足（相入玉に対局者同士が合意し，点数計算で 24 点未満となった.）
  - 被入玉宣言（相手が条件を満たした状態で入玉を宣言した.）

- 以下の状態になった場合には，引き分けとなる．

  - 連続王手以外の千日手（連続王手以外で同一局面が 4 回現れた.）
  - 持将棋（相入玉に両プレイヤが合意し，点数計算で両者ともに 24 点以上となった．図 2.2 は持将棋の局面例である.）

将棋は，ゲーム理論の分類ではおおむね「二人零和有限確定完全情報ゲーム」とされる．同じ分類の似たゲームとして，世界で人気のあるゲームとしてチェスが挙げられる．将棋とチェスの違いは，将棋のほうが 9×9 とチェスの 8×8 の盤面よりも広い，駒の種類が 8 種類とチェスの 6 種類より多い，チェスにはない「持ち駒」が存在するという理由から局面がより複雑になりやすい．また，局面における平均合法手数もチェスよりも多い．

将棋は終局に近づいても駒の数の総和は等しいままで，局面が複雑な状態で切り合いになって終わることが多い．これに対して，他のゲーム，例えばチェスでは終局に近づくにつれて駒が少なくなり，なかなか詰まない場合もある．囲碁では戦いはおおむね中盤で起こり，終盤は互いの陣地の境界を定めて静かに終わる

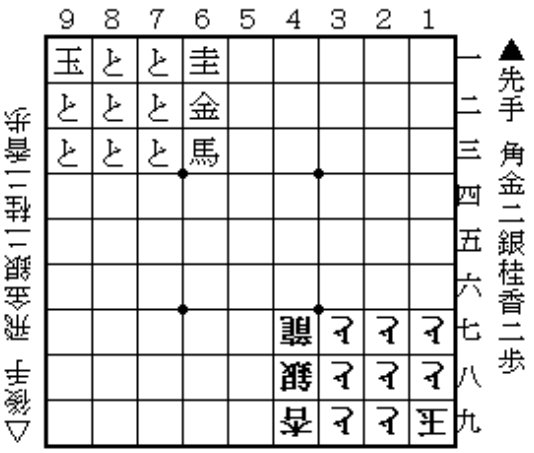


図 2.2: 持将棋の局面例．この局面では先手後手ともに 27 点あり，24 点以上なため持将棋で引き分けになる．

ことが多い．このように終局の仕方が異なることは他のゲームの既存研究を応用する際には注意すべき点である．

2.2 持ち時間

将棋の対局における持ち時間設定は大きく分けて，切れ負け，秒読み，フィッシャールールがある．切れ負けは決められた持ち時間を使い切ると即時間切れ負けになる持ち時間設定である．秒読みは，持ち時間を使い終わった後も一手ごとに定められた時間内に指せば時間切れにならないという持ち時間設定である．また，秒読みの時間に加えて 1 局を通して定められた回数・時間を使用できる考慮時間も採用されている場合がある．フィッシャールールは定められた持ち時間に加えて，一手ごとに決められた時間が加算されていく持ち時間設定である．ここではアマチュアが主に使用する将棋倶楽部 24，将棋クエスト，将棋ウォーズという 3 つの対局サービスで採用されている持ち時間設定について述べる．

将棋倶楽部 24 では，以下の持ち時間設定が採用されている．

- 早指し 1：持ち時間 1 分，切れたら 1 手 30 秒
- 早指し 2：持ち時間なしで 1 手 30 秒に加えて 60 秒の考慮時間
- 早指し 3：持ち時間 5 分，1 手ごとに 5 秒加算
- 15 分：持ち時間 15 分，切れたら 1 手 60 秒
- 長考：持ち時間 30 分，切れたら 1 手 60 秒

将棋クエストでは切れ負けが採用されており，将棋クエストでは 2 分切れ負け・5 分切れ負け・10 分切れ負けが選択できる．将棋ウォーズでは切れ負けと秒読みルールが採用されており，切れ負けでは 3 分切れ負けと 10 分切れ負けが，秒読みでは 10 秒将棋が選択できる．

2.3　対局の段階と戦略

将棋の対局は大きく分けて「序盤」「中盤」「終盤」の 3 つの段階がある．序盤は初手から攻めの陣形や守りの陣形（囲い）が完成して駒がぶつかり合うまで，中盤は序盤が終わってからどちらかのプレイヤの囲いが崩れ始めるまで，終盤は中盤が終わってから終局するまで，というふうに言われている．ただし，序盤・中盤・終盤の境目は曖昧であり，人によって意見が異なることもある．

一つの対局を序盤・中盤・終盤に分けると，各段階ごとに目標がある．序盤では効率良く攻めの陣形や囲いを構築することが目標であるが，攻めの陣形や囲いの強さは相対的なものであり，相手の指し手と相談しながら決めることが大切である．序盤では，陣形を整備する手が目につきやすい．中盤では駒得（相手の駒を取ったり，自分の価値が低い駒と相手の価値が高い駒を交換したりすること）や，自分の攻め駒と相手の守り駒を交換すること，自分の駒を成ることを目標とする．ただし，相手も同じことを目指しているので，自分が得をすることだけでなく，損をしないことについての考慮も大切である．中盤では，駒を取る手は目につきやすく，逆に駒を取られる手は指しづらい．また，次に取られそうな駒を逃がす手もよく指される．終盤では，基本的に相手玉の詰みを目指して攻めたり，自玉の詰みを防ぐために受けたりすることが目標になる．将棋ではチェスと異なり持ち駒を利用でき，終盤になっても駒の総数は変わらず，合法手の数も多い．そのため，終盤になっても局面が単純にならず，初心者同士では終局させること自体に苦労することもある．終盤では，駒の損得が比較的重要ではなくなり，詰みの有無や，玉が受けなしになるかどうかが重要になる．これらは似た形でも結論が異なることもあるため，先読みを正確に行うことが大切である．終盤では，王手が最も目につきやすい．

第3章　関連研究

ゲームにおいてAIは，対戦相手として使用されているだけではなく，仲間として人間プレイヤをアシストしたり，教師として人間プレイヤの成長を補助したり，解説者として観戦者の理解を助けたりするといった用途で使用されている．まず簡単に対戦相手以外の例を示しておくと，仲間AIに関して，佐藤らは協力型RPGにおいて人間プレイヤの好みを推定することで，人間プレイヤの不満を低減することに成功した[11]．教師AIに関して，池田らは囲碁において人間プレイヤの上達を助けるため，人間プレイヤの悪手を検出し，ラベル付けする研究を行った[12]．この研究は囲碁棋士によって明らかに有用なレベルであると評価された．解説AIに関して，亀甲らは将棋において棋力の低い観戦者の理解を助けるため，与えられた局面を表現する特徴的な単語を予測し，予測した単語の情報と言語モデルを組み合わせることで解説文を生成する研究を行った[13]．また，AIは人間がゲームをプレイするときだけでなく，バランス調整[14]やProcedural Content Generation[15]のようにゲームのコンテンツの調整・生成を行うときにも活用される．

続いて対戦相手としてのゲームAIに話を戻すと，あまりにも弱すぎると多くの人間を楽しませることは難しいため，一定以上の強さが求められる．ゲームAIを作るための古くから用いられている単純なアプローチの一つとして，ルールベース[1]が挙げられる．ルールベースは,「条件」と「条件が満たされたときの行動」で構成されるルールに基づいてゲームAIの行動を決定する手法である．ルールベースの弱点としては，ゲームAI開発者がそのゲームに習熟している必要があるうえ，習熟している知識について細かな場合分けが必要であり，開発者の能力に大きく依存するということが挙げられる．また，複雑なゲームにおいて，ルールベースアプローチのみでは人間を超えるような性能の実現はほとんど不可能である[1]．ゲームAIを実現する他の手法としては，手作りの局面評価関数と$\alpha\beta$法のような探索手法を組み合わせるというアプローチも挙げられる．この手法は一定の成果を収め，チェッカーなどのゲームでは人間のトッププレイヤと同じレベルのゲームAIが作られた[16]．

ここまでのヒューリスティックな手法では，比較的複雑な将棋や囲碁といったゲームにおいては人間のプロに勝つことは難しかった．そこで，局面評価関数を作りやすい将棋のようなゲームでは，棋譜からの学習により自動的に局面評価関数を作る手法が現れた[17]，また，局面評価関数が作りにくい囲碁のようなゲームでは，シミュレーション結果を利用して有望な手を深く読むモンテカルロ木探索（MCTS）という手法が現れた[18]．この手法にも着手の確率分布を得るために人

間の棋譜からの学習が行われている．このようなデータからの学習による局面や着手の評価関数の作成や，シミュレーションを用いた優れた探索手法を用いることで，ゲーム AI は大幅に強化され，少なくとも将棋や囲碁においてはアマチュア有段者レベルの棋力を手に入れた．そして，人間にとってありえないミスをしにくくなる，というような意味合いで人間らしさも改善された．

ここまで紹介したゲーム AI の手法では，人間の思考や学習法や実際の挙動を模倣しようとしていた．しかし，ビデオゲームでは人間のプレイデータを使用しない深層強化学習によって一般的な人間プレイヤのパフォーマンスを上回り [19]，将棋や囲碁，チェスでは人間の棋譜を使用せずに自己対局から得た棋譜から学習を行う AlphaZero というゲーム AI が各ゲームでの世界チャンピオン AI を破った [20]．特に，AlphaZero については人間のチャンピオンを超える強さをも手に入れたため，この時点から強さと人間らしさが明確に異なる方向を向いたのだと推測できる．

ではこのような人間よりも強いゲーム AI が存在する場合，人間らしさについて向上させる必要はあるのだろうか．これについて，Mcllroy-Young らは，AI がある分野で人間の性能を超えると，人間がその分野のタスクを AI に任せる場合と，人間が AI と協調してその分野のタスクに取り組む場合があると述べている [4]．ゲーム AI の分野は後者に属し，このような分野では適切に設計された AI は人間に対して支援したり，指導したり，適切な情報を与えたりする機会が豊富にある．しかし，強いチェス AI や探索深さを浅くするといった方法で弱体化させたチェス AI は人間らしくないため，解釈が難しく，AI から学習することが困難であるとも指摘している [4]．このような理由から，我々は人間らしいゲーム AI を実現する価値が高いと考える．

ゲーム AI の人間らしさを評価する方法として，多くのゲームで利用できるチューリングテストがある．ゲーム AI についてのチューリングテストでは，評価者は対戦している相手もしくは観察しているプレイヤが人間かゲーム AI なのかを判定する．ゲーム AI は人間と誤認された割合が人間らしさの評価になる．ゲーム AI についての代表的なチューリングテストの競技会については，一人称視点シューティングゲーム（FPS）では 2K BotPrize，横スクロールアクションゲームでは Mario AI/Platformer AI Competition が挙げられる．前者の競技会では，MirrorBot というゲーム AI は，少しの遅延を挟んで対面した相手の動きを模倣するという手法で平均的人間プレイヤよりも人間らしいと評価された [6]．後者については，競技会で使用されたものではないが，藤井らの「疲れ」や「遅れ」などの生物学的制約を課した強化学習によって構成されたマリオ AI が，人間プレイヤよりも人間らしいと実験によって評価された [21]．

将棋においては，生井らは，棋譜が多数入手できるプレイヤの棋風を模倣するため、対象者の棋譜のみから作った定跡データベースを用いて戦略を，対象者の棋譜を標準の局面評価関数に追加で学習することで局面評価を模倣することを目指した [7]．仲道は，局面評価関数や探索深さに手を加えて弱体化させた将棋 AI 同士の棋譜と，人間の棋譜について，人間らしいか，AI らしいかを判断させる実験

を行った [5]．その結果，判断する者の棋力が高いほど悪手に対する感度が高くなり，不自然さを感じやすくなることを示した．

第4章　人間らしい価値関数

本章は，第 26 回ゲームプログラミングワークショップで発表した,『対局状況をより正確に表現するための盤面評価値』（発表論文リスト [1]）という論文の内容をもとに再構成したものである．本章では，局面から勝率を予測する価値関数について，棋力情報を直接入力に含めることで，より実際に近い人間の勝率予測や，逆転が起こりやすい局面の自動抽出についての研究を行う．

4.1　背景と目的

将棋は，強いゲーム AI から楽しませるゲーム AI へという流れの中で，良い題材としてさまざまな観点から取り上げられてきた．池田は，楽しませるゲーム AI の要素技術として，プレイヤを理解したり，不自然な着手を抑制したりすることが重要だと述べている [22]．また，仲道は，評価者の棋力と感じる不自然さには大きな関わりがあると指摘している [5]．仲道はさらに，ゲーム AI との対局は人間との対局に比べてつまらないという問題点を指摘している．例えば将棋であれば，将棋 AI と調整を行わずに戦うと，棋力が高すぎて対局相手として適しておらず，探索を調整すると人間とは異なった弱さが現れ不自然に感じる．このように，対局相手として将棋 AI を評価すると，残された課題は多い．

対局以外については，対局の解説や棋譜の解析において将棋 AI に人間らしさが求められる場合がある．例えば，対局を観戦しているときは，どのような展開になるか，勝敗はどうなりそうかといった内容を把握できたほうが対局の臨場感が味わえ，より楽しむことができる．ただし，これらの情報を局面のみから把握するためには，対局者以上の棋力が必要になることも多い．ここで活躍するのが人間の「解説者」や「聞き手」である．最近では，解説に将棋 AI が用いられたり，人間がおらず将棋 AI の局面評価関数（価値関数）のみが参考にされたりする場合がある．前者のような将棋 AI を用いた解説では,「AI の推奨する手は人間には指せない，人間ならばこう指すところ」といったプロによる指摘がよく見受けられる．このような指摘は，既存の将棋 AI が最善のやりとりを探索した結果の局面評価値と候補手を表示しており，人間にとって高度すぎる手順を示しているときに見られる．この場合，示されている評価値はそれ以降も AI によって正確な着手選択が行われることを前提としたものなので，人間プレイヤの実感や実際の対局結果を予測できるように表せていない可能性がある．

この課題はプロ棋士の対局に限ったことではなく，アマチュアが将棋を勉強する際にも問題になる．アマチュアが将棋を指す場合，プロ棋士よりも時間が短く，切れ負けや最初から秒読みの将棋が多い．短時間で納得のいく局面まで読み切ることは難しいので，制限時間に従って読みを入れることになる．その場合だと，最善手を指し続ければ必ず勝つが一手でも間違えれば即負けにつながる，という局面評価値が高い局面よりも，必ず勝てるわけではないが何度かミスをしても勝てそうな，少し局面評価値が低い局面を選ぶほうが，結果的に勝率が高くなることが多い．そのため，対局者の強さに応じた「実感・実際の結果に近い」局面評価関数を求めることの価値は高いと考える．

本研究の目的は，対局者同士の平均棋力を考慮することで，より実感・実際に近い状況と表現できる人間らしい局面評価関数を設計することである．

4.2 関連研究

中屋敷らは，既に完全解析されていて後手必勝であることが判明しているどうぶつしょうぎ [23] を題材にして，局面評価関数に逆転の余地を取り入れることで容易な勝ちと難しい勝ちを区別し，完全解析の結果よりも人間にとって役立つ局面評価関数を設計した [24]．この研究では，方策のエントロピーと逆転のしやすさを結び付けることで，劣勢である先手番での逆転勝ちや，優勢である後手番での安全勝ちを狙うことに優れた局面評価関数を得ている．

Maia[4] は，AlphaZero[20] 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膨大な量の人間の棋譜から深層教師あり学習しているチェス AI 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として知られている．この研究では，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，探索を行わずに人間の着手予測や，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．Maia の研究では，対局者のレートが 1000 から 1100 なら Maia 1000 が学習する，という方法をとっているため，大量の棋譜が必要になる，レートを連続値として扱えない，という問題がある．

4.3 提案手法と考えられる用途

既存の将棋 AI の多くは，強い将棋 AI 同士が指し続けた場合どのような結果になるか，という探索に基づいて局面の評価を求めている．これは，すでに述べたように，一般的な人間プレイヤの実感や，人間プレイヤが実際に指した場合の結果と異なる場合がある．そこで我々は，局面とプレイヤの棋力から，その棋力帯のプレイヤたちが指し続けた場合の勝率を求めたい．

本研究では局面と棋力情報を入力として勝敗を出力とするような教師あり学習で近似することを目指す．例えば,「相手玉に難解な詰みがあるが，自玉は明らか

な必至である」という局面においては，棋力が低ければ勝率が低く，棋力が高ければ勝率が高いと判定されるようにしたい．提案するモデルの利点は，学習する際にレートを連続値として扱えるため，比較的必要なデータ数が少量で良いこと，また，実際に使用する際に使用者のレートにより近づけられることが挙げられる．

もし本研究により，局面とプレイヤの棋力から，その棋力帯のプレイヤたちが指し続けた場合の勝率が予測できた場合は，以下のような用途が考えられる．

- プロ棋士の対局において，予想される結果に近い，実際の形勢を示すことができる．
- 初中級者の指導 AI として，無理のないアドバイス，実質的に勝率を上げるためのアドバイスが可能になる．
- 一度ミスすれば形勢が極端に悪化するのか，それとも何度かミスが許されるのかという，局面の安定度のようなものに利用できる[1]．
- 初中級者の対局相手 AI として，安定度と人間らしい方策関数を組み合わせることで，調整なしでは生じにくい攻め合いへの誘導を達成することができる．

4.4 実験設定

実験では，将棋倶楽部 24 万局集 [26] を使用した．9 割の棋譜を訓練データとし，1 割の棋譜をテストデータとして用いた．その際，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，対局者のレートの差が 50 以内の棋譜を抽出した．レート差が 50 というのは上位者の勝率が 57%，下位者の勝率が 43%という実力差になる．

また，最序盤と入玉含みの局面を除外するため，手数が 50 手目から 199 手目までの局面を使用した．最序盤を除外した理由は，棋士の最善手率などを計算・比較する研究 [27] でも時代により変遷の多い定跡部分について分けて分析することに倣った．入玉含みの局面を除外した理由は，将棋倶楽部 24 万局集が発行された時点では将棋倶楽部 24 に入玉宣言法が実装されておらず，入玉の適切な評価を行えないためである．その結果，訓練データは約 686 万局面，テストデータは約 76 万局面が得られた．

レートを直接入力に含めることが有効かを確認するため，レートなし条件とレートあり条件という 2 つの条件を比較する．レートなし条件では局面のみを入力として，レートあり条件では局面と対局者二人のレートを平均したものを入力とした．入力に使用したレートは平均が 0，分散が 1 になるように標準化を行っている．レートなし条件，レートあり条件ともに局面の勝率と着手確率分布を同じネットワークで予測するマルチタスク学習を行っている．

[1] 安定度に似た概念として，共謀数が挙げられる [25]

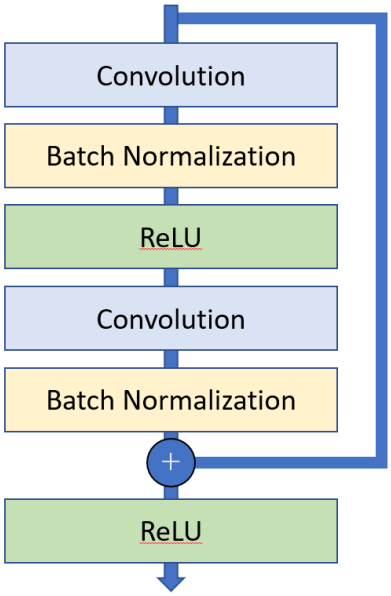


図 4.1: ResNet の構成

ネットワークの構造や学習オプションについては，python-dlshogi ライブラリ[2] を参考にした．入力の形式は駒ごとの配置の配列，手番ごとの占有している座標の配列，手番ごとの持ち駒の配列によって構成されているビットボードから入力特徴を作成した．入力特徴は駒の種類ごとにチャンネルを分け，各チャンネルは駒の座標を表す二値画像とした．持ち駒は種類ごとに最大枚数分のチャンネルを割り当てた．中間層のフィルター枚数は 256 として，1 層目は畳み込み層，2 層目からは図 4.1 のような Residual Network(以下 ResNet) を 5 ブロック重ねる形を採用した．出力の形式については，勝率予測では，勝敗を二値分類の問題として捉え，着手確率分布予測では，移動方向と移動先の座標の組み合わせにより，2187 のラベルを分類する多クラス分類の問題として捉えた．オプティマイザは Adam を採用した．

[2]https://github.com/TadaoYamaoka/python-dlshogi 最終確認日 2023 年 2 月 1 日

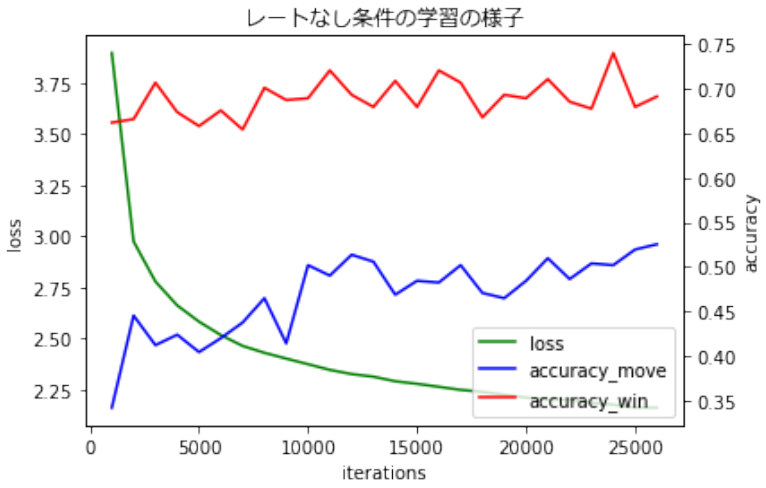


図 4.2: レートなし条件での学習の様子

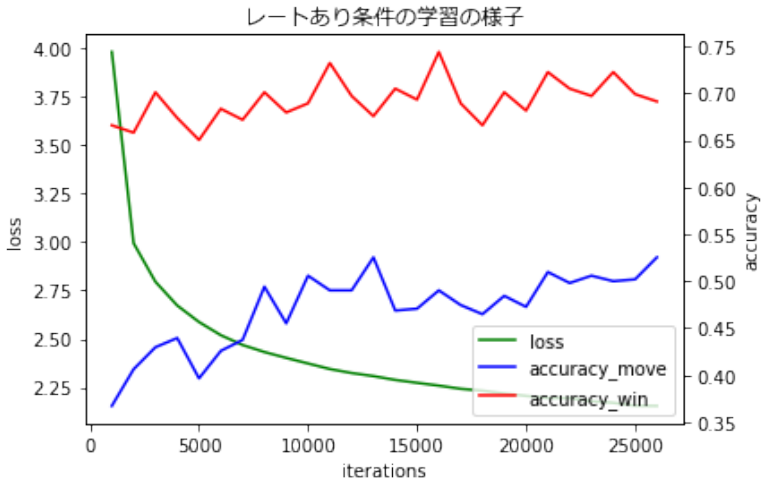


図 4.3: レートあり条件での学習の様子

4.5 学習結果の統計量

図 4.2 は，レートなし条件での学習の様子，図 4.3 は，レートあり条件での学習の様子であり，training データについての loss と accuracy を示したものである．学習曲線はおよそ 25000iteration で収束したとみなした．図 4.2 と図 4.3 を比較すると，レートなし条件でもレートあり条件でも学習の挙動は概ね同じだといえる．

学習にかかった時間は，GTX-1070 の一般的な PC でおよそ 4 時間だった．このため，学習する棋譜が数倍程度に増えても問題ない．学習には約 12 万棋譜を使用したが，これは Maia の 1 モデルにつき使用した 1200 万棋譜と比較すると約 100 分の 1 の数である．

表 4.1: 手数ごとの予測勝率の正解率の比較

手数	50-79	80-109	110-139	140-169	170-199
局面数	363112	265826	98299	26388	6183
レートなし条件	0.6459	0.7393	0.7658	0.7831	0.7847
レートあり条件	0.6457	0.7395	0.7676	0.7889	0.7962

表 4.2: 平均レートごとの予測勝率の正解率の比較

平均レート	0-499	-999	-1499	-1999	-2499	2500-
局面数	81885	208233	184283	200948	83516	862
レートなし条件	0.6942	0.7135	0.7038	0.6907	0.6857	0.7135
レートあり条件	0.6947	0.7132	0.7060	0.6906	0.6857	0.7135

実験の結果，勝率予測の正解率は，レートなし条件では 0.7000，レートあり条件で 0.7005，着手予測の正解率は，レートなし条件では 0.4977，レートあり条件でははは 0.4983 という結果になった．着手予測，勝率予測のどちらもレートあり条件のほうが正解率が高いが，有意な差ではなかった．これらの条件に大きな差が見られないことは意外な結果であった．我々は，棋力によって結果が異なる局面が多くあると考えており，レートを入力に入れることで，予測がより正確になると推測していたためである．結果から，この前提が間違っている可能性も示唆されるが，別の可能性もある．局面は対局者が選んだ手の積み重ねであるため,「初心者らしい局面」や「プロらしい局面」といったものが存在する．そのため，レートを明示的に入力に含めずとも局面から棋力の大まかな推定が出来てしまい，推定精度の差がほとんど出なかったという可能性がある．

ここからは勝率予測を分析していく．表 4.1 は，レートあり条件での予測勝率の正解率と，レートなし条件での予測勝率の正解率を，手数ごとにまとめたものである．結果から，どちらの条件でも手数が進んで終局に近づくほど勝率の予測が正確になることが分かった．また，終盤に近づくにつれて，レートあり条件とレートなし条件の正解率の差が大きくなっている．このことから，終盤に近いほど，レートを入力に含めて学習することでより状況を正確に判断できることが分かった．これは，先ほど述べたような，難解かつ勝敗に直結するような手が終盤だと現れやすいからだと考えている．表 4.2 は，レートあり条件での予測勝率の正解率と，レートなし条件での予測勝率の正解率を，対局者二人の平均レートごとにまとめたものである．平均レートの高低によってレートあり条件とレートなし条件での正解率に差は出なかった．また，どちらの条件でも，レートが高いほど予測がしやすいというわけでもないことが分かった．

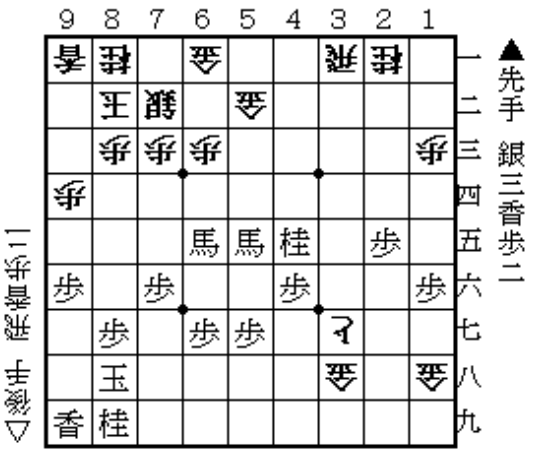


図 4.4: 的確な攻めが必要な局面

4.6 具体的な局面を用いた評価

予測勝率が実際の勝率に近いかという評価については，人間によって指し継ぎ（途中局面からの対局）を行うことが理想的だが，現実的には困難なため，ある程度弱くした将棋 AI によって指し継がせる．予測勝率の評価のため，形勢判断に人間的な項目を採用している技巧 2 によって指し継ぎを行う．Bonanza という将棋 AI は，仲道らによってレートが算出されており [5]，探索の深さが 4，5 の Bonanza の強さはそれぞれ，将棋倶楽部 24 のレートで 1740，1984 に相当する．探索の深さが 1 の技巧 2 は，探索の深さが 4 の Bonanza に 61 勝 38 敗 1 分，探索の深さが 5 の Bonanza に 42 勝 58 敗 0 分であったので，将棋倶楽部 24 のレートで 1850 程度として使用した．

4.6.1 的確な攻めが必要な局面

図 4.4 は指し継ぎを行った 1 つ目の局面で，先手番である．この局面を水匠 4 で 10 億ノード読ませると，▲2 一馬△ 4 八飛▲ 7 八香△ 3 四飛▲ 9 五歩という手順で先手優勢 (1260) と評価する．評価値を勝率に変換する式については，パラメータ a を使用して，1 / (1 + exp(-評価値/a))[28] とされている．本研究では，a についてよく使用される値である 600 を採用する．この式に 1260 という評価値を当てはめて勝率に変換すると，先手勝率 0.8909 となる．つまり，先手優勢ではあるが的確な攻めを行う技量が求められる局面ということである．

一方，本研究のモデルに評価させると，対局者のレートが 1850 のときは，先手の勝率が 0.2283 と評価する．すなわち，このレート帯アマ三段程度では攻めに失敗することが多いだろうと推測されている．

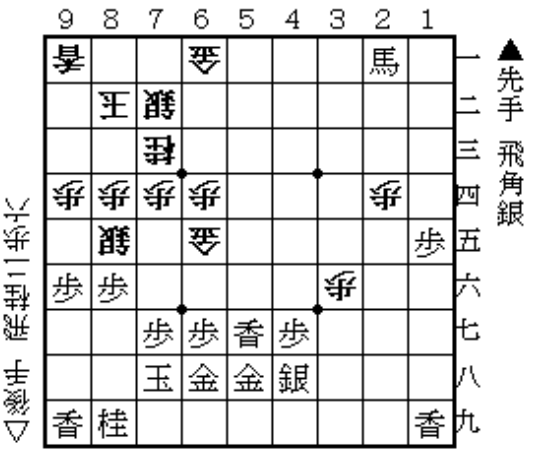


図 4.5: 的確な受けが必要な局面

実際に，探索の深さが 1 の技巧 2 同士で指し継がせると，先手 35 勝後手 65 勝引き分け 0 と，本研究のモデルの評価に比較的近い結果となり，これは推測がうまくいっている例である．

4.6.2 的確な受けが必要な局面

図 4.5 は指し継ぎを行った 2 つ目の局面で，後手番である．この局面を水匠 4 で 10 億ノード読ませると，△ 8 六銀▲ 4 三馬△ 8 七桂▲ 8 八銀△ 9 九桂成という手順で先手有利 (760) と評価する．これを勝率に変換すると先手勝率 0.7802 となる．つまり，先手優勢ではあるが的確な受けが求められる局面ということである．

一方，本研究のモデルに評価させると，対局者のレートが 1850 のときは，先手の勝率が 0.4314 と評価する．すなわち，このレート帯（アマ三段程度）では受けに失敗することがあり，互角に近いだろうと推測されている．

実際に，探索の深さが 1 の技巧 2 同士で指し継がせると，先手 50 勝後手 50 勝引き分け 0 と，本研究のモデルの評価に比較的近い結果となり，これも推測がうまくいっている例である．これらの指し継ぎによって，少なくともある局面においては，本研究のモデルが既存のモデルよりも対局の状況をより実際に近く表現できたといえる．

4.6.3 予測勝率差が大きい局面

我々は，入力レートによって予測勝率が大きく異なる局面について解釈を与えることで本稿のモデルについて一定の評価を下せると考えた．60 手目から 79 手目，80 手目から 99 手目，100 手目から 119 手目より，棋譜が被らないようにそれぞれ

図 4.6: 予測勝率差が最も大きい局面 (60〜79 手目)

512 局面ランダムサンプリングした．それぞれの手数のサンプルについてレートが 500 の場合とレートが 1850 の場合で勝率予測を行い，予測勝率の差が最も大きい局面を抽出した．図 4.6 から図 4.8 が実際の局面図である．なお，先後にかかわらず，手番が全て手前側に来るように調整した．後に述べる勝率も手番側にとっての勝率である．以下で局面の解説と解釈を行う．

図 4.6 の局面では，レート 500 が入力された場合は勝率 0.5996，レート 1850 が入力された場合は勝率 0.3065 と出力される．後手は桂馬と香車を得しており，先手からの攻めをいなせば有利になりそうな局面である．先手からは 4 四歩という攻めが目に付く．5 三金とあたりを避けると同銀成同角 4 三歩成がある．後手が困っているようだが，実は 4 四歩には同金と取る手が成立し，同角 4 三香と打って反撃すれば後手が指せる局面である．レート 500 同士の対局であれば先述の攻めが成功するが，レート 1850 同士の対局になると比較的多くの人間が反撃の筋に気づくために予測勝率の差が大きくなるのだと解釈できる．

図 4.7 の局面ではレート 500 が入力された場合は勝率 0.3496，レート 1850 が入力された場合は勝率 0.6313 と出力される．先手が王手をかけられている局面で，先手が正しく応対すれば勝てそうだが，頓死筋があったり，後手玉が絶対に王手がかからない局面であったりするため，逆転が起こりやすい．具体的には，8 五玉で詰まないと思って指すと 8 六金で頓死したり，7 六玉と指した後に 9 八竜と切って同香に 6 四金と抑える筋があり，実戦的には少し大変である．ただ，ある程度の棋力があれば受けきれるために，レート 1850 が入力された場合は先手有利になるのだと解釈できる．

図 4.8 の局面ではレート 500 が入力された場合は勝率 0.7839，レート 1850 が入力された場合は勝率 0.5097 と出力される．先手の銀と桂馬と後手の角が交換され，後手のほうが駒得だが玉が薄い．ここでは 2 六香が最善手で，4 四桂は手順前後で互角になる．具体的には，4 四桂同歩 2 六香同飛車と切って 2 一香と打つ手があ

図 4.7: 予測勝率差が最も大きい局面 (80〜99 手目)

図 4.8: 予測勝率差が最も大きい局面 (100〜119 手目)

る．4 四桂を指していなければこのとき 2 四歩が打て，特に問題ない．このように後手の反撃筋は飛車を切って小駒で攻めをつなげるような難易度の高いものになり，初級者同士であれば先手がそのまま押し切れると考えたため，低レートを入力した場合は先手勝率が高くなり，高レートを入力した場合は予測勝率が互角に近くなるのだと考えた．

第5章　人間らしい方策関数

本章は，第27回ゲームプログラミングワークショップで発表した,『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』（発表論文リスト [2]）という論文と，第15回International Conference on Agents and Artificial Intelligence(ICAART) で発表した『Improving the Human-Likeness of Game AI ' s Moves by Combining Multiple Prediction Models』（発表論文リスト [3]）という論文の内容をもとに再構成したものである．本章では，局面から着手確率を予測する方策関数について，チェスで最も有効な手法として知られている深層教師あり学習が将棋でも有効かを確認する．また，AlphaZero 的な方策についても，人間の手をうまく予測できるかを確認し，各方策の長所と短所を明らかにする．特に，深層教師あり学習の方策については，モデルが予測しづらい局面を考察・分類することで，より理解を深める．そして，深層教師あり学習の方策と AlphaZero 的な方策を組み合わせることで，人間らしさを向上させる 2 つの手法を提案する．

5.1　背景と目的

人間らしい方策関数に関して，Mcllroy-Young らは，従来の強いゲーム AI や，それを弱体化したものの方策は，チェスにおいて 30%から 40%程度しか人間の手を予測できず，人間の方策とは大きく異なると述べている [4]．これは対局に勝つことだけが目的であれば問題ないが，人間プレイヤと対局して楽しませることを目的とする場合には問題になる．なぜなら，人間プレイヤはゲーム AI の手が不自然である，もしくは理解できないと感じると，対局を楽しむのは難しいからである．

人間の手を再現するだけであれば，着手確率最大の手を出力すれば良いが，もし棋力帯ごとの，もしくは特定の人間プレイヤが指す任意の手の自然さが人間に見分けがつかないほどの精度で予測できれば，様々な応用が考えられる．例として，詰将棋の難易度を評価したり，教師 AI としてプレイヤにとって最も勝ちやすい手を示したりすることができる．

本研究の目的は，強みと弱みが異なる複数の方策をうまく組み合わせることで，より人間らしい方策関数を設計することである．

5.2 関連研究

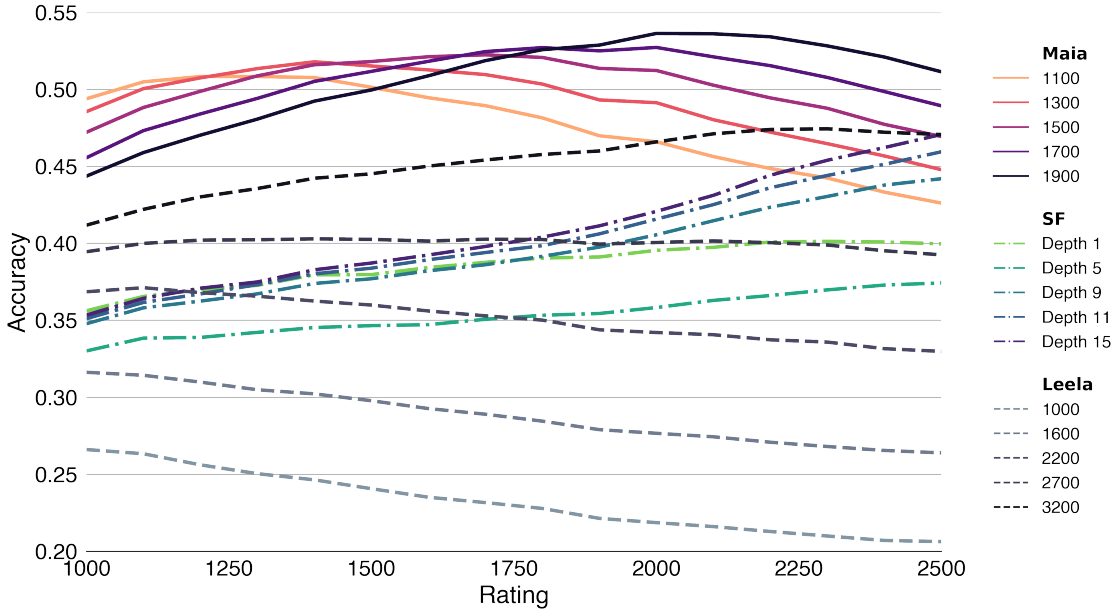


図 5.1: Maia，Stockfish（SF），Leela の一致率．[4] より引用．

Maia[4] は，AlphaZero[20] 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深層教師あり学習しているチェス AI である．チェスにおいては，人間の着手を予測する最も効果的な研究の一つとして知られている．この研究では，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行う際に，人間プレイヤの棋譜をレートごとに 9 つのグループに分割し，各グループについて 1200 万棋譜を用いて着手を予測するモデルの学習を行った．図 5.1 は Maia，MinMax 系のチェス AI（Stockfish；SF），AlphaZero 系のチェス AI（Leela）の人間の棋譜との一致率を表したものである．その結果，Maia は一致率が 50%から 55%程度と，MinMax 系のチェス AI や AlphaZero 系のチェス AI よりも人間プレイヤの着手を予測できることが分かった．また，MinMax 系のチェス AI や AlphaZero 系のチェス AI は，おおむね強いもののほうが一致率が高く，それらの AI は強いプレイヤとの一致率ほど高かった．

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 [29]．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．

杵淵らは，直前の手の流れを考慮することで将棋 AI の一致率を向上させる手法を提案した [8]．この研究では，幅広い棋力を持つプレイヤを対象とし，探索をもとにした局面評価関数 [17] と遷移確率関数 [30] を組み合わせることで手の流れを

表現した．線形結合が用いられ，その重みは人間の着手から学習することで決められた．杵淵らの提案した手法は各々の関数単独よりも人間の着手を予測できた．

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 [31]．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．

我々は，これらの関連研究の結果を踏まえて，強みが異なる複数の方策を組み合わせることで，人間らしさについても向上させられるのではないかと考えた．

5.3 提案手法

Maia の結果は有望な結果であったが，我々は，さらに議論する価値のある 2 つの問題を見つけた．評価指標と，少量のデータを用いた場合の人間らしさの改善の余地である．評価指標については，関連研究については一致率，すなわち AI が最も高い確率で予測した着手が人間プレイヤの着手に一致したか，という指標がよく用いられる [4][29]．この指標は，人間を模倣する能力を評価するにあたって，ある程度有効な指標である．しかしながら，一致率には，予測と異なる手が，人間にとって自然な手であっても，ありえないほど不自然な手であっても，等しく不一致と評価してしまうという弱点がある．つまり，一致率が高いからといってそのゲーム AI が全体的に人間らしいとは言えず，一致しなかった手についてはまったく人間らしくない手を指してしまうことがある．

人間らしさを評価する別の評価指標として，我々は尤度もまた妥当な指標だと考えた．尤度とは，モデルが人間の着手をどれくらいの確率で予測していたか，という指標である．尤度の積はモデルの方策と人間の方策が等しいときのみ最大になる．したがって，尤度は人間の方策を模倣するという目標に沿った評価指標といえる．

次に少量のデータを用いた場合の人間らしさの改善の余地について述べる．Maia はチェスを題材に高い精度の方策関数の学習に成功したが，使用された棋譜の枚数は一つのレート帯あたり 1200 万棋譜である．これは将棋・囲碁・麻雀といった他のメジャーなゲームですら収集が難しい規模であり，より一般的なゲームでの利用を考慮すれば，比較的少数の棋譜でもある程度の精度の方策関数を得られるような方法が望まれる．このような場合，一致率や尤度は，探索によって改善するかもしれない [29] し，AlphaZero[20] のような強化学習系の強いゲーム AI の方策を利用することで改善するかもしれない．

本研究の概要は以下のとおりである．まず，将棋の場合，Maia のような教師あり学習が人間の手をうまく予測できるかどうかを，一致率と尤度の 2 つの指標で確認する．また，教師あり学習の方策と AlphaZero 的な方策を比較して，各方策の長所と短所を明らかにする．特に，教師あり学習の方策については，尤度の低

い局面を考察・分類することで，より理解を深める．そして，教師あり学習の方策と AlphaZero 的な方策を組み合わせることで，一致率，尤度を向上させる 2 つの手法を提案する．

本研究では Maia の手法に従い，教師あり学習にニューラルネットワークを使用する．K 個のクラスを持つ多クラス分類に用いられるニューラルネットワークを考え，入力を x，出力の一つを $u_k(-\infty < u_k < \infty)$ とする．x がクラス C_k に属する確率 $p(C_k|x)$ はよく次のように表され，

$$p(C_k|x) = \frac{exp(u_k)}{\sum_{j=1}^{K} exp(u_j)} \tag{5.1}$$

x は $\arg\max_k p(C_k|x)$ のクラスに分類される．将棋の場合では，ある局面 x に対し，合法手が K 個あれば，方策関数は着手 $C_k(k = 1$ から $K)$ の確率 $p(C_k|x)$ から作られる．

一致率は $\arg\max_k p(C_k|x)$ が実際の着手と一致するか，という指標であり，人間の着手を予測する研究でよく利用されている．しかし，この一致率という指標は確率分布を正しく評価できないという弱点がある．例として，60%の人間が a，30%の人間が b，10%の人間が c の手を指すとし，この時，a: b: c = 90%: 5%: 5%とする予測モデルと，34%: 33%: 33%とする予測モデルの 2 つがあるとする．どちらのモデルでも着手 a の予測確率が最も高いため，一致率は 60%となる．ゲーム AI の強さを向上させるために人間を模倣していた時代はどちらのモデルでも問題なかった．しかし，人間らしさのために人間を模倣する場合，分布の形が重要になる．理想的な確率分布は，人間がよく指す手には高い確率を，あまり指さない手には低い確率を与える，つまり人間と同じ形状の分布である．

この問題を解決するため，ある方策が人間の着手をどれだけ予測できるかを評価する別の指標として，尤度を用いる．この研究では，局面 x の集合 X と，それに対応する人間の着手 C_{human} が与えられたとき，以下のように尤度を計算する．

$$(\prod_{x \in X} p(C_{human}|x))^{\frac{1}{|X|}}, \tag{5.2}$$

ここで，$|X|$ は X のサイズ，$p(C_{human}|x)$ は方策から C_{human} の確率を予測したものである．ここで計算した尤度は，別の言い方をすれば，方策によって予測された確率の幾何平均といえる．尤度は，人間の方策とモデルの方策が等しいときのみ最大となるため，人間の方策を模倣するための合理的な指標である．本研究では，一致率と尤度の両方を用いてモデルの人間らしさを評価する．

5.3.1 Classifier モデル

複数の方策を組み合わせる一つ目の手法は分類器（Classifier）を用いたモデルである．図 5.2 に Classifier モデルの概要を示した．人間の着手予測するために，

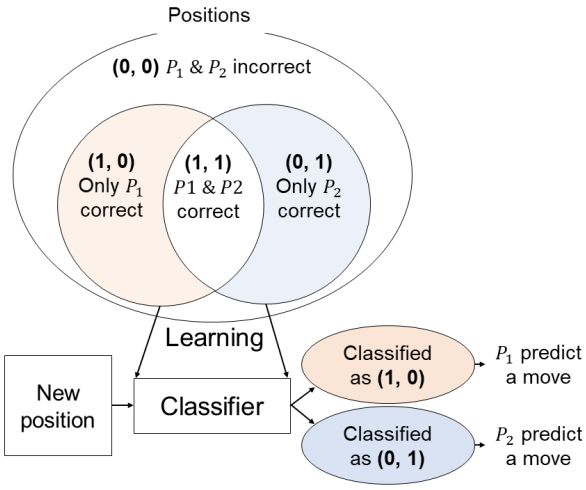


図 5.2: Classifier モデルの概要．(0, 0) は P_1 と P_2 両方とも人間の着手を正しく予測できない局面，(1, 0) は P_1 のみが人間の着手を正しく予測できる局面，(0, 1) は P_2 のみが人間の着手を正しく予測できる局面，(1, 1) は P_1 と P_2 の両方が人間の着手を正しく予測できる局面を意味する．

それぞれ異なる長所と短所を持つ 2 種類の方策 P_1 と P_2 があるとする．我々は，P_1 に適した局面では P_1 で予測し，P_2 に適した局面では P_2 で予測したい．これを実現するために，P_1 に適した局面か，P_2 に適した局面かを分類器を使って判断する．

分類器の学習データを作成するために，局面と人間の着手のセットを用意し，P_1 と P_2 に各局面の人間の着手を予測させる．局面は，P_1 のみが人間の着手を正しく予測した場合，(1, 0) とラベル付けされ，P_2 のみが人間の着手を正しく予測した場合 (0, 1) とラベル付けされる．そして，これらの（局面，ラベル）の組を用いて教師あり学習により分類器を学習させる．新規の局面で人間の着手を予測するとき，局面が (1, 0) に分類される場合は P_1 が，(0, 1) に分類される場合は P_2 が使用される．この分類器を用いると，各局面に対して比較的適切な方策を用いることができる．この手法は，2 値分類だけでなく，3 つ以上の方策から最も適切な方策を選択するマルチクラス分類や，適切な方策をすべて選択するマルチラベル分類に拡張することができる．

5.3.2 Blend モデル

複数の方策を組み合わせる 2 つ目の手法は，方策の値を混ぜ合わせる（Blend）モデルである．複数の学習器の推定結果を統合することで，より良い精度を得る機械学習アルゴリズムであるアンサンブル学習にヒントを得て，異なる方策からの確率を混ぜ合わせる手法を提案する．

Classifier モデルと同様に，長所と短所が異なる 2 つの方策があるとする．p_{1k} を方策 P_1 から得られた着手 k の確率，p_{2k} を方策 P_2 から得られた着手 k の確率とし，

$\alpha(0 \leq \alpha \leq 1)$ を P_1 の重みを決定するパラメータとする．新しい確率 p_k は以下のように計算される．

$$p_{new\,k} = p_{1k}^{\alpha} \times p_{2k}^{(1-\alpha)} \tag{5.3}$$

$$p_k = \frac{p_{new\,k}}{\sum_{j=1}^{K} p_{new\,j}} \tag{5.4}$$

この式は 2 つの確率を非線形に混ぜ合わせているが，

$$p_{new\,k} = \alpha \times p_{1k} + (1-\alpha) \times p_{2k} \tag{5.5}$$

というふうに線形に混ぜ合わせる式や，

$$p_{new\,k} = (\alpha \times p_{1k}^{\beta} + (1-\alpha) \times p_{2k}^{\beta})^{1/\beta} \tag{5.6}$$

という (5.3) や (5.5) を拡張した式もありうる．本研究では，予備実験の結果から，式 (5.3) は (5.5) より一致率や尤度が高く，(5.6) と同程度の性能だったため，シンプルで高性能な式 (5.3) を使用する．

5.4 将棋における着手予測実験

本節では，Maia がチェスで行ったような教師あり学習が，将棋においても高い精度で人間の方策を予測できるか，また，AlphaZero 的な方策がどの程度人間の手を予測できるかを確認するための実験を行った．人間の手の予測精度を評価する際には，一致率と尤度の 2 つの指標を用いた．

5.4.1 実験設定

将棋において人間の指し手を予測するために，Maia の研究と同様に方策関数と価値関数の深層教師あり学習を行った．Mindwalk 株式会社から提供を受けた将棋クエスト[1]の 10 分切れ負け 300 万局を使用し，以下の 3 種類の不適切とみなしたデータを除外した．まず，時間切れで負けた対局は除外した．これは，最も操作しやすい駒を動かすなど，時間切れ負けに特有の行動が考えられるためである．次に，プレイヤのレート差が 50 以上の対局を除外した．これは，レーティング差があると，強いプレイヤが極端に不利な状況から勝利する可能性があり，価値関数の学習に悪影響を与える可能性があるためである．最後に，手数が 50 手目以降の局面を使用した．序盤の局面を除外した理由は，棋士の最善手率などを計算・比較する研究 [27] でも時代により変遷の多い定跡部分について分けて分析することに倣った．

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．

[1] http://questgames.net 最終確認日 2023 年 2 月 1 日

- グループ 1：R1433〜R1591
- グループ 2：R1592〜R1655
- グループ 3：R1656〜R1708
- グループ 4：R1709〜R1768
- グループ 5：R1769〜R1855
- グループ 6：R1856〜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．結果として，各モデルの学習には 13 万棋譜を使用した．これは，Maia の 1200 万棋譜と比較すると，約 100 分の 1 のデータセット数である．

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，python-dlshogi2 ライブラリ[2]を参考にした．ライブラリと大きく異なる部分は，入力に現在の局面だけでなく，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．本研究でも予備実験で直近 12 局面を入力に含めると，現局面のみから着手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．

学習は各群について，10 エポックずつ行った．これは，10 エポック程度で loss が収束することが多いことを予備実験で確認したためである．学習には，RTX-3070 GPU を搭載した PC で，各群ごとに約 4 時間かかった．

議論を簡単にするため，これらの Maia のようなモデルを Maia-S(S は small data, Shogi の頭文字) と表記し，グループ 1 を用いて学習したモデルを Maia-S-1，グループ 2 を用いて学習したモデルを Maia-S-2，以下同様に表記する．

Maia-S モデルのほかに，AlphaZero ベースの将棋 AI である，DLshogi[3]を比較対象として採用した．DLshogi は，2022 年に開催された第 32 回世界コンピュータ将棋選手権で優勝したプログラムである．本論文では，“DLshogi policy” は DLshogi の prior（探索を行わない状態での方策）を，“DLshogi visits” は DLshogi の MCTS によって得られた訪問回数に対応した着手確率分布を意味する．DLshogi visits を使用する理由は，AlphaZero[20] の手法において，prior の学習に MCTS によって得られた訪問回数に対応した着手確率分布を使用しているためである．実験では，DLshogi の MCTS のノード数を 10,000 に制限した．

[2]https://github.com/TadaoYamaoka/python-dlshogi2 最終確認日
[3]https://github.com/TadaoYamaoka/DeepLearningShogi 最終確認日 2023 年 2 月 1 日

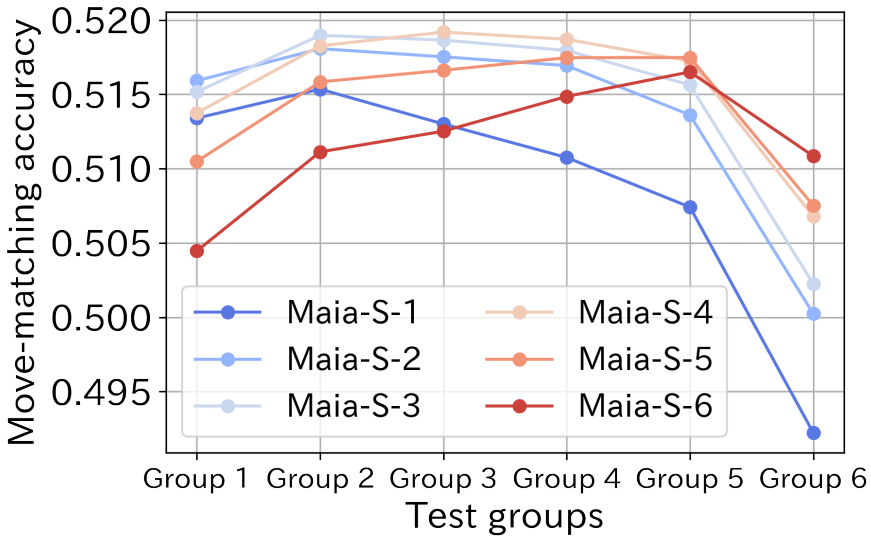


図 5.3: Maia-S モデルの一致率

5.4.2 実験結果

本節では，まず Maia-S の結果について議論する．図 5.3 は，グループごとにテストした Maia-S モデルの一致率を示している．すべてのグループにおいて，そのグループを最もよく予測するモデルの一致率は 51.0%から 52.0%であった．例えば，グループ 1 では，Maia-S-2 の一致率 0.515 が Maia-S モデルの中で最も優れていた．将棋においても Maia の手法である程度一致率を高めることができることが確認された．異なるゲームなため，単純な比較はできないが，この精度はチェスにおける Maia に近い．一般的な傾向として，予測モデルの学習データのレートとテストデータのレートが近ければ予測性能はよく，遠ければ予測性能が悪くなることが判明した．例えば，グループ 1 については Maia-S-2 が，グループ 2 については Maia-S-3 が，グループ 3 については Maia-S-4 が，グループ 4 から 6 については，対応した Maia-S-4 から Maia-S-6 が最も予測性能が高い．

次に，AlphaZero 的な方策が人間の手をうまく予測できるかを分析した．DLshogi policy と DLshogi visits の結果を，それぞれ図 5.4 の黄色の曲線と黒色の曲線で示す．どちらの方策も，レートの高い人間の手をより正確に予測できる傾向があった．特に，DLshogi policy のグループ 6 に対しての性能は Maia-S-6 を上回っている．この傾向は，Jacob らの Maia のモデルと探索を組み合わせて使用するとき，探索の効果はレートに依存しているという主張 [29] にも合致している．DLshogi の結果と比較すると，Maia-S モデルの一致率はグループ間で 1 ポイント程度の差しかない．これらの結果から，Maia-S モデルはレートによらず高い精度で人間の手を予測でき，DLshogi policy はレートの高い人間の手をより正確に予測できると結論づけられる．

また，一致率に加え，尤度についても分析を行った．我々は，Maia-S モデルと

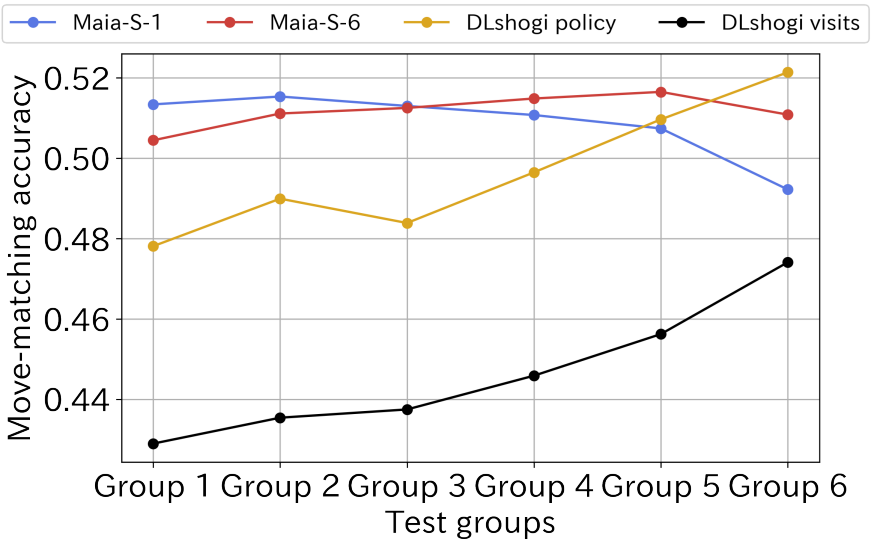
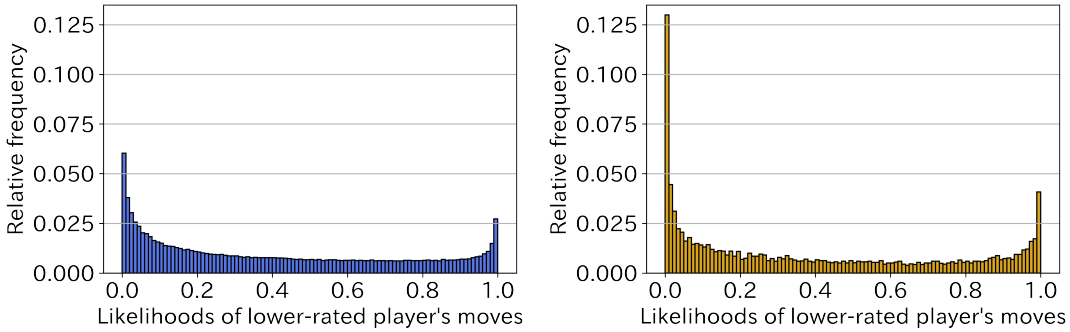


図 5.4: Maia-S モデルと DLshogi の一致率



(a) Maia-S-1 モデル　　(b) DLshogi policy

図 5.5: 低レートプレイヤの尤度のヒストグラム

DLshogi policy の性能差が最も大きい低レートプレイヤのデータ（グループ 1）に着目した．図 5.5 は Maia-S-1 モデルと DLshogi policy における低レートプレイヤデータの尤度のヒストグラムである．x 軸は尤度を 100 ビン（0.00-0.01, 0.01-0.02, ... , 0.99-1.00）に分割したもので，y 軸は各ビンの相対頻度である．どちらの分布も二峰性で，両端に 2 つの山がある．具体的には，0.00-0.01 のビン付近に山があり，このビンにある着手はモデルが予測しづらいことを意味する．もう一つの山は 0.99-1.00 のビン付近にあり，このビンにある着手はモデルが予測しやすいことを意味する．Maia-S-1 モデルでは 6%程度，DLshogi policy では 13%程度の頻度で尤度が 0.00-0.01 の手が存在する．ここから，モデルが予測できていない人間の手が少なからず存在することが分かる．

本節では，Maia-S モデルと DLshogi policy は人間の着手予測においてある程度高い一致率があり，それぞれに長所があることを示した．また，Maia-S，DLshogi

policy はともに尤度が低い手が存在することを示した．次節では，尤度が低い手や局面について考察・分類を行う．

5.4.3　学習モデルが予測しづらい局面の考察・分類

尤度のヒストグラムには 2 つの山があったが，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予測精度を改善するという立場からは興味の対象ではない．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 つに分類した．

- M1．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．このように筋の良いように見える手があって探索しないとそれが悪手と分からない場合は，最善手はそれと比較して有望な手には見えず，予測確率が低くなってしまう．モデルが探索しないことによるミスについては，探索が必要な局面かを予測し，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．

- M2．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．限られた枚数の学習だと，周囲が似た状態でこのような見落としの棋譜を学習しているケースは少ないので，そうした人間の悪手は予測確率が低くなってもおかしくない．このような人間らしいミスを再現する（今よりも高い確率で起きうるとみなす）ためには，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，という方法がありうる．

- M3．操作ミス
  操作ミスについては，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．

- M4．様々な手が有望な局面
  様々な手が有望な局面では，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うえ，実際に指される手は一つなので，一致率も低くなる．

- M5．敗勢の局面
  どうしよう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おり，本質的には様々な手が有望な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．

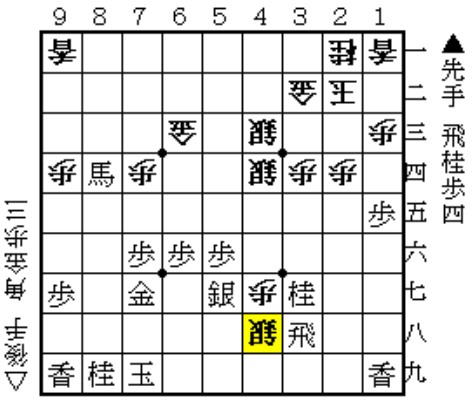


図 5.6: モデルが探索をしないことが原因の局面
先手番
予測勝率：0.37
予測上位 5 手：[▲ 4 八同銀（0.430），▲ 2 三歩（0.187），▲ 6 八銀（0.187），▲ 6 一飛（0.036），▲ 4 八飛（0.035）]
実際に指された手：▲ 6 五歩（0.00123）

分類した局面の中で，典型的なものを取り上げる．また，分析のために MinMax 系で最も強い将棋 AI の一つである水匠を使用した．図 5.6 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．

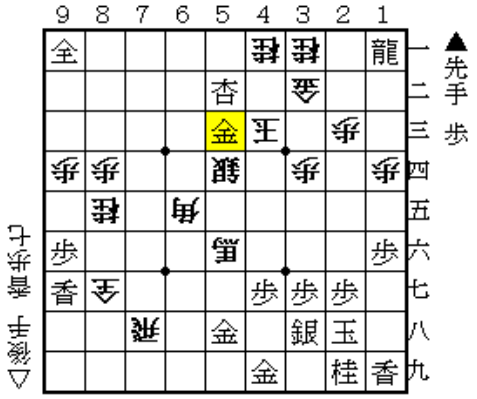


図 5.7: 人間の探索に関するミスが原因の局面
後手番
予測勝率：0.489
予測上位 5 手 [△ 5 三桂（0.968），△ 4 四玉（0.020），△ 3 三玉（0.005），△ 5 三玉（0.005），△ 5 三銀（0.000）]
実際に指された手：△ 3 三玉（0.00532）

図 5.7 は，人間の探索に関するミスを原因に分類した局面である．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 三玉は互角（局面評価値：-213，水匠 5 で 1000 万ノード探索）である．場合にもよるが，このような手を再現できれば，将棋 AI が自然に逆転されるようになり，人間プレイヤをより楽しませることができると考えている．

図 5.8: 操作ミスが原因の局面
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 5.8 は，操作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操作ミスで 9 八飛成を選択している．これは予測しづらい人間らしいミスと言えるが，このようなミスを再現しても大多数の人間プレイヤに喜ばれるとは思えないので再現する必要性は薄いと考えている．

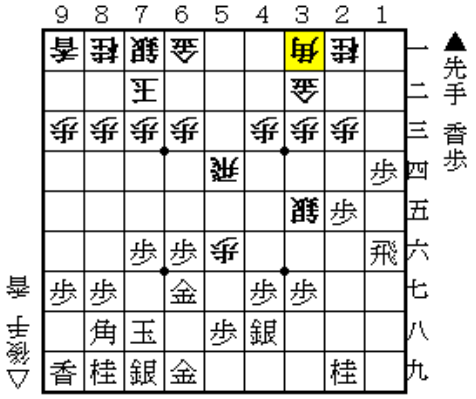


図 5.9: 様々な手が有望な局面
先手番
予測勝率：0.633
予測上位 5 手：[▲3 六歩（0.221），▲6 五歩（0.216），▲1 三歩成（0.136），▲6 八銀（0.084），▲2 八香（0.045）]
実際に指された手：▲7 七角（0.00406）

図 5.9 は様々な手が有望なことを原因に分類した局面である．3 一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，自玉を固める手など，手の選択肢の幅が広い．実際に指された手は 7 七角と自玉を固める手で自然ではあるものの，様々な手が有望であるため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）と様々な手が有効である．また，予測上位 5 手の確率も最大で 0.221，最小で 0.045 とばらついている．

図 5.10: 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）

図 5.10 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，プレイヤは他のどんな手を指しても結局負けるので，様々な手がありうる．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．

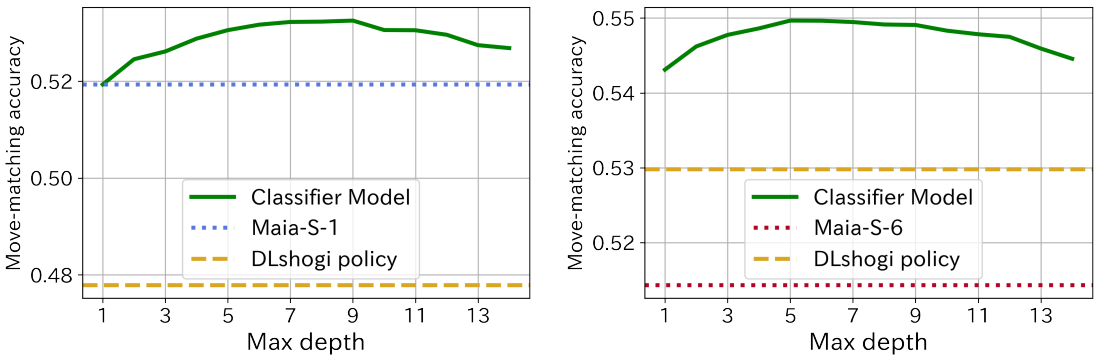
5.5 Classifer モデル

この節では，5.3.1 項で提案した Classifier モデルについて，一致率と尤度の向上に関する評価実験を行う．

5.5.1 実験設定

5.4 節の実験の低レート（グループ 1）と高レート（グループ 6）のテストデータから 45,000 局面をランダムサンプリングした．分類器として，Python の Scikit-learn ライブラリのランダムフォレスト[4]を採用し，max depth の設定を変更する以外はデフォルトのパラメータ設定のまま使用した．また，担当を決めるための分類器への入力として，Maia-S による局面 x に対する $\max_k p(C_k|x)$，DLshogi policy に

[4]https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html 最終確認日 2023 年 2 月 1 日



(a) 低レートプレイヤのデータ　　(b) 高レートプレイヤのデータ

図 5.11: Classifier モデル，Maia-S モデル，DLshogi policy の一致率

よる局面 x に対する $\max_k p(C_k|x)$，Maia-S と DLshogi policy の KL 情報量（確率分布間の距離のようなもの）という 3 つの特徴量を使用した．$\max_k p(C_k|x)$ を特徴量として採用した理由は，多クラス分類においてはこの値が推論の信頼性として解釈できるので，適切な方策選択に役立つと考えたためである．KL 情報量を特徴量として採用した理由は，モデル間の距離を測ることで，近い場合はどちらのモデルでもあまり違いはないが，遠い場合は適したモデルとそうでないモデルを選ぶことに大きな違いがあるという情報を与えることで，推論の信頼性向上に役立つと考えたためである．分類器は (1, 0) または (0, 1) を出力し，与えられた局面に対して Maia-S と DLshogi policy のどちらがより適切であるかを決定する．評価には 10-fold cross-validation を用い，各テストデータに対する一致率の平均を計算した．

5.5.2　実験結果

図 5.11 は max depth を変更したときの Classifier モデルの一致率のグラフである．また，比較のため，Maia-S モデルと DLshogi policy の結果も載せた．

低レートプレイヤのデータについて，Maia-S-1 モデルと比較して，max depth が 9 のときに一致率が 0.519 から 0.533 に改善した．また，高レートプレイヤのデータについて，DLshogi policy と比較して，max depth が 5 のときに一致率が 0.530 から 0.550 に改善した．ここから，低レートプレイヤ，高レートプレイヤについて Classifier モデルが元のモデル単体よりも高い一致率を得ていることが分かる．この結果から，Classifier モデルを用いて，より適切な方策を選択することで，精度を向上させられることが分かった．

また，各モデルの尤度の比較も行った．低レートプレイヤの場合，Maia-S-1 モデルの尤度は 0.196，DLshogi policy の尤度は 0.129，Classifier モデルの尤度は 0.169 となった．高レートプレイヤの場合，Maia-S-6 の尤度は 0.198，DLshogi policy の

尤度は 0.188，Classifier モデルの尤度は 0.190 であった．つまり，Classifier モデルでは，一致率は向上するものの尤度は向上しない．これは一致率だけでは人間らしさの完全な指標にはなりえないという一例と言える．今回の特徴量はシンプルな特徴量のみを採用したが，より複雑な特徴量を採用することで性能を改善できる可能性がある．また，この手法ではどのような分類器についても採用できるため，さまざまな種類の分類器について結果を調べれば，ランダムフォレストよりも性能が良い分類器が存在する可能性がある．

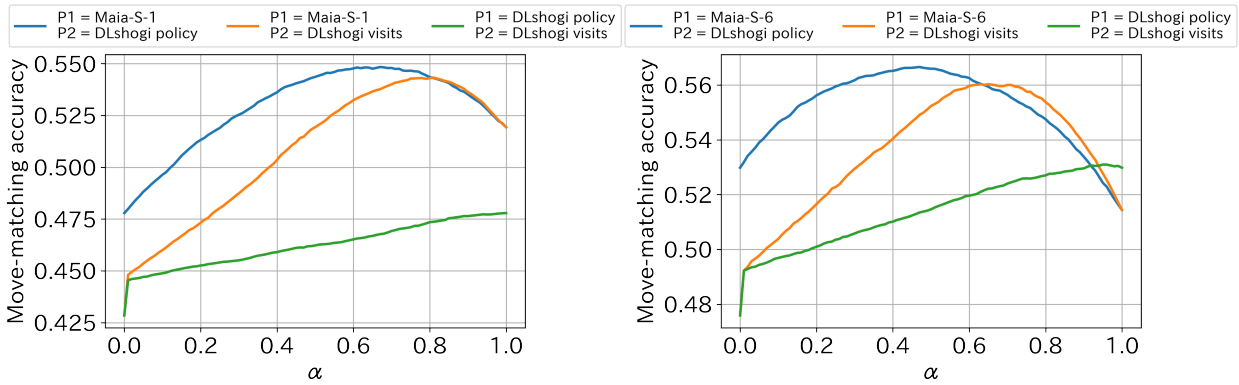
5.6 Blend モデル

この節では，5.3.2 項で提案した Blend モデルについて，一致率と尤度の向上に関する評価実験を行う．

5.6.1 実験設定

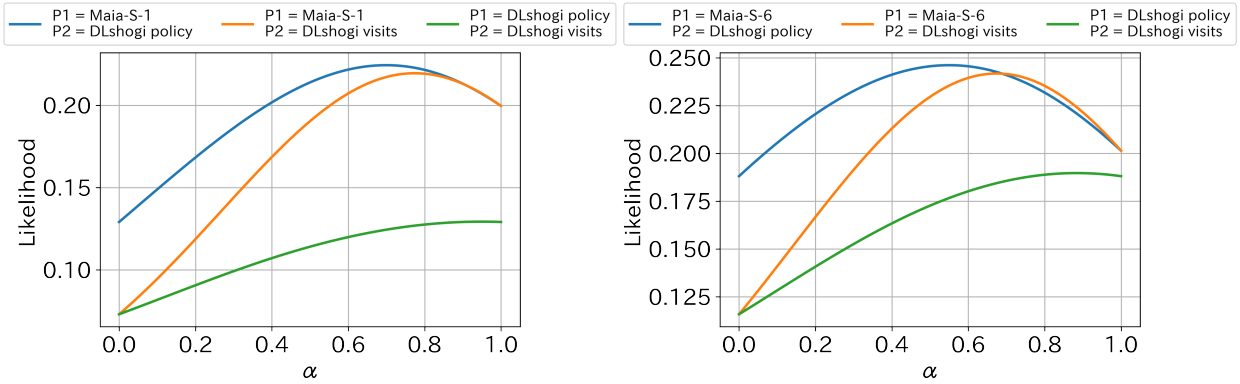
5.1 節と同じく，グループ 1 とグループ 6 の 45,000 局面を使用して，Blend モデルの検証を行った．モデルは $p_k = p_{1k}^{\alpha} \times p_{2k}^{(1-\alpha)}$ を使用する．P_1，P_2 の候補は，Maia-S，DLshogi policy，DLshogi visits である．DLshogi visits を含むことが，グループ内の全局面ではなく 45,000 局面を使用した主な理由であり，DLshogi visits は 1 局面ごとの確率分布を得るのに約 4 秒かかる．

5.6.2 実験結果



(a) 低レートプレイヤのデータ　　(b) 高レートプレイヤのデータ

図 5.12: α と P_1，P_2 の組み合わせを変化させた場合の Blend モデルの一致率



(a) 低レートプレイヤのデータ　　(b) 高レートプレイヤのデータ

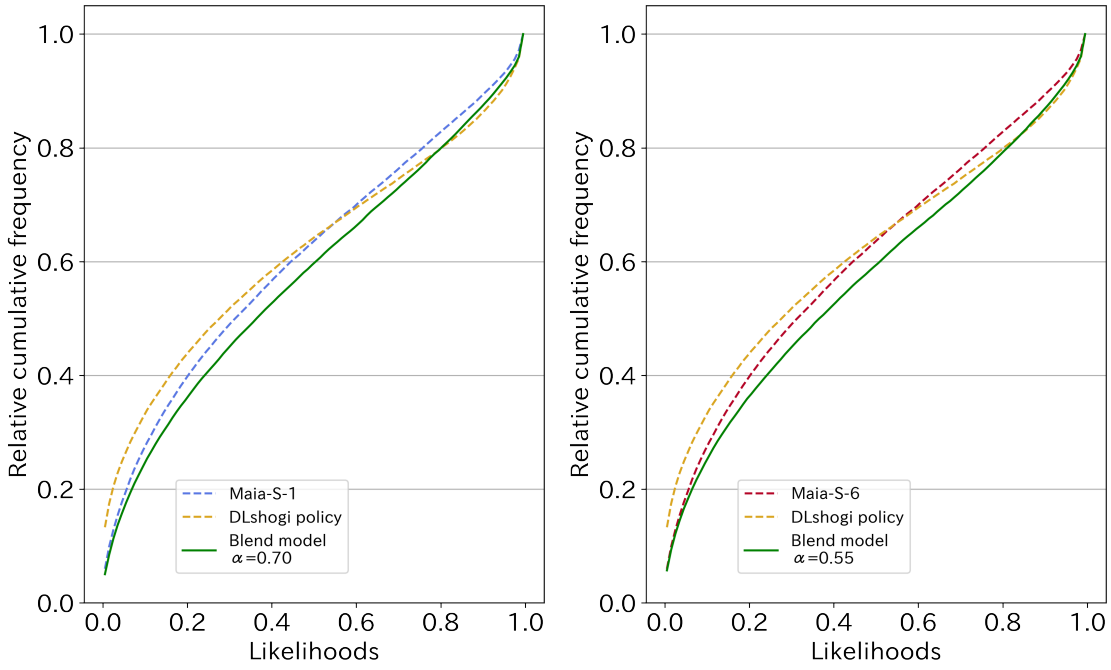
図 5.13: α と P_1，P_2 の組み合わせを変化させた場合の Blend モデルの尤度

図 5.12，図 5.13 は α を 0.0 から 1.0 まで変化させた場合の Blend モデルの一致率，尤度である．その結果，Maia-S と DLshogi policy の組み合わせ（青い曲線）から，低レートプレイヤのデータでも高レートプレイヤのデータでも，他の組み合わせより高い一致率と尤度を得た．DLshogi policy と DLshogi visits の組み合わせ（緑の曲線）については，曲線は単調に増加する傾向を示したが，一致率や尤度は最高でも DLshogi policy 単独と同程度（$\alpha = 1.0$）であり，組み合わせる価値が低いことが分かった．この理由について，DLshogi policy と DLshogi visits が非常に似た方策であるために価値が低くなったのだと推測できる．

Blend モデルでは，Maia-S モデルと DLshogi policy を組み合わせ，最適な α を設定した場合，低レートプレイヤのデータでは 0.519 から 0.548，高レートプレイヤのデータでは 0.530 から 0.567 に一致率を向上させることができた．また，尤度に関しても低レートプレイヤのデータでは 0.200 から 0.224 に，高レートプレイヤのデータでは 0.201 から 0.246 に改善した．

Blend モデルの α の値は P1 の重み，例えば青い曲線にある Maia-S の重みを表す．図 5.12(a) と 5.12(b) を比較すると，P1 の重みである α の最適値は高レートプレイヤのデータよりも低レートプレイヤのデータのほうが高い傾向があることが分かった．5.4 節では Maia-S モデルが比較的低レートプレイヤのデータを予測するのに適しており，DLshogi policy が比較的高レートプレイヤのデータを予測するのに適していることを示した．α の傾向もこの結果と合致している．

図 5.14 は Blend モデル，Maia-S モデル，DLshogi policy の尤度の相対累積頻度である．尤度が 0 に近いほど人間の手の予測確率が低く，尤度が 1 に近いほど人間の手の予測確率が高い．Blend モデルの曲線（緑色の実線）は，Maia-S モデルや DLshogi policy の曲線を概ね下回っており，Blend モデルは Maia-S モデルや DLshogi policy よりも高い確率で人間の手を予測したことが分かる．



(a) 低レートプレイヤのデータ　　(b) 高レートプレイヤのデータ

図 5.14: Blend モデル，Maia-S モデル，DLshogi policy の尤度の相対累積頻度

第6章　おわりに

本研究では，限られた枚数の棋譜から価値関数と方策関数を学習する場合でも，より人間らしくする手法について提案を行った．

人間らしい価値関数については，人間同士の対戦を前提として，実際・実感により近い勝率を予測することを目標に，教師あり学習の入力にレートを含めて勝率を推定した．モデルの予測勝率が実際の勝率に近いかの評価を行うため，弱い将棋 AI による指し継ぎを行った．結果として，弱い将棋 AI の指し継いだ勝率について，強い将棋 AI の予測勝率と比べ，提案したモデルの予測勝率のほうが近かった．また，同一局面で入力する棋力を変えた場合に，予測勝率が大きく異なる局面をサンプリングして，局面の解釈を行った．その結果，このサンプリング方法で抽出したそれぞれの局面は，たしかに逆転が起こりやすい局面であろうことを確認できた．

将来的には，うっかりや勘違いなどのミスを再現した探索モデルとともに使用することで，ミスが起こりやすい局面の人間的な評価も目指したい．また現時点では，指し継ぎにおいて将棋 AI は十手程度の詰みがあればほとんど詰ませてしまう．アマチュアや，持ち時間が短いときのプロが詰みを見逃すことはよくあるので，人間らしい終盤モデルや，人間らしい終盤評価関数の実現を目指したい．

人間らしい方策関数についてはまず，チェスで人間の着手を予測する研究で最も有効な手法の一つとして知られている Maia の深層教師あり学習アプローチが将棋でも有効かを確かめた．本論文ではこのモデルを Maia-S と名付け，将棋においても人間の着手予測に有効であることを示した．また，自己対局から学習する AlphaZero 的なモデルが人間の手をどれくらいの精度で予測できるかについても分析を行った．その結果，AlphaZero 的なモデルは，より棋力の高いプレイヤの手をより正確に予測することが分かった．そして，モデルについてより理解を深めるため，尤度（モデルによって予測される人間の手の確率）に着目して，モデルが予測しづらい局面について考察・分類を行った．

これらの分析に基づき，複数の方策を組み合わせることで，人間らしさをさらに向上させる手法を提案した．1 つ目のアプローチは，Classifier モデルという，分類器を用いて各局面でより適した方策で人間の手を予測するものである．2 つ目のアプローチは，Blend モデルという，異なる方策が出力する確率をブレンドするものである．前者の手法では，一致率を 1 ポイントから 3 ポイント向上させることができたが，尤度の改善は見られなかった．また，後者の手法では，2 ポイントから 5 ポイント程度の精度の向上が見られ，尤度についても 2 ポイントほど改善した．

今後の課題として，いくつかの方向性がある．現在 Maia-S モデルによる探索は行っていない．この Maia-S モデルと木探索を，Jacob らの手法のように組み合わせて，将棋において人間の手の予測精度を向上させられるかを分析する予定である．また，別の方向性として，モデルが予測しづらい局面や人間の手を分析し，それを再現できる新しい手法を取り入れることでも精度の改善が期待できる．また，尤度がどの程度，その着手が選択されやすいかを反映しているかを調べることも重要であろう．

謝辞

本論文は，筆者が北陸先端科学技術大学院大学先端科学技術研究科池田心研究室に在籍中の成果をまとめたものです．研究を進めるにあたり，池田心教授には多大な支援と指導を賜りました．躓きがちな自分を辛抱強く見守り，温かく励ましてくださったおかげで，安心して研究を進めることができました．心より感謝いたします．Hsueh Chu-Hsuan 助教には，特に国際論文の執筆に関して，大変丁寧な指導をしていただきました．また，困っているときには常に最大限の支援を申し出ていただいたこと，とても心強かったです．深く感謝申し上げます．Mindwalk 株式会社の棚瀬寧氏には，貴重な研究データである，将棋クエストの棋譜を提供していただきました．誠にありがとうございます．

研究室の皆様や同学の友人には，研究だけでなく日常生活においても多くのことを助けてもらいました．おかげで 2 年間楽しく研究生活を送ることができました．

最後に，常にあたたかく応援してくれた家族に心から感謝します．

発表論文リスト

[1] 小川竜欣, 池田心. 対局状況をより正確に表現するための盤面評価値, 第 26 回ゲームプログラミングワークショップ (GPW), pp.28–33, (2021).

[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, 第 27 回ゲームプログラミングワークショップ (GPW), pp.180–186, (2022).

[3] Ogawa, T., Hsueh, C.-H., Ikeda, K.: Improving the Human-Likeness of Game AI ' s Moves by Combining Multiple Prediction Models, 15th International Conference on Agents and Artificial Intelligence (ICAART), Paper #276, (2023)

参考文献

[1] 伊藤毅志, 保木邦仁, 三宅陽一郎：ゲーム情報学概論―ゲームを切り拓く人工知能―, コロナ社 (2018).

[2] Campbell, M., Hoane Jr, A. J. and Hsu, F.-h.: Deep blue, *Artificial intelligence*, Vol. 134, No. 1-2, pp. 57–83 (2002).

[3] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al.: Mastering the game of go without human knowledge, *nature*, Vol. 550, No. 7676, pp. 354–359 (2017).

[4] McIlroy-Young, R., Sen, S., Kleinberg, J. and Anderson, A.: Aligning superhuman AI with human behavior: Chess as a model system, in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1677–1687 (2020).

[5] 仲道隆史：熟達度に着目した将棋プログラムに対する不自然さに関する研究, PhD thesis, 電気通信大学 (2020).

[6] Polceanu, M.: MirrorBot: Using human-inspired mirroring behavior to pass a turing test, in *2013 IEEE Conference on Computational Inteligence in Games (CIG)*, pp. 1–8 (2013).

[7] 生井智司, 伊藤毅志：将棋における棋風を感じさせる AI の試作, 研究報告ゲーム情報学 (GI), Vol. 2010, No. 3, pp. 1–7 (2010).

[8] Kinebuchi, T. and Ito, T.: Shogi Program That Selects Natural Moves by Considering the Flow of Preceding Moves, in *2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence*, pp. 79–84 (2015).

[9] 将棋 [https://ja.wikipedia.org/wiki/%E5%B0%86%E6%A3%8B] (最終確認日 2023 年 2 月 1 日).

[10] 持ち時間 [https://ja.wikipedia.org/wiki/%E6%8C%81%E3%81%A1%E6%99%82%E9%96%93#%E5%B0%86%E6%A3%8B] (最終確認日 2023 年 2 月 1 日).

[11] Sato, N., Ikeda, K. and Wada, T.: Estimation of player's preference for cooperative RPGs using multi-strategy Monte-Carlo method, in *2015 IEEE Conference on Computational Intelligence and Games (CIG)*, pp. 51–59 (2015).

[12] Ikeda, K., Viennot, S. and Sato, N.: Detection and labeling of bad moves for coaching go, in *2016 IEEE Conference on Computational Intelligence and Games (CIG)*, pp. 1–8 (2016).

[13] 亀甲博貴, 三輪誠, 鶴岡慶雅, 森信介, 近山隆：対数線形言語モデルを用いた将棋解説文の自動生成, 情報処理学会論文誌, Vol. 55, No. 11, pp. 2431–2440 (2014).

[14] 高木幸一郎, 雨宮真人：ロールプレイングゲーム（RPG）のバランスとは何か：分析およびその調整に関する提案, Technical Report 58(2001-GI-006), 九州大学大学院システム情報科学研究科, 九州大学大学院システム情報科学研究科 (2001).

[15] Nam, S.-G., Hsueh, C.-H. and Ikeda, K.: Generation of Game Stages With Quality and Diversity by Reinforcement Learning in Turn-Based RPG, *IEEE Transactions on Games*, Vol. 14, No. 3, pp. 488–501 (2022).

[16] Schaeffer, J., Lake, R., Lu, P. and Bryant, M.: CHINOOK: The World Man-Machine Checkers Champion, *AI Mag.*, Vol. 17, pp. 21–29 (1996).

[17] Hoki, K. and Kaneko, T.: Large-scale optimization for evaluation functions with minimax search, *Journal of Artificial Intelligence Research*, Vol. 49, pp. 527–568 (2014).

[18] Coulom, R.: Computing "Elo ratings" of move patterns in the game of Go, *ICGA journal*, Vol. 30, No. 4, pp. 198–208 (2007).

[19] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. and Riedmiller, M.: Playing atari with deep reinforcement learning, *arXiv preprint arXiv:1312.5602* (2013).

[20] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K. and Hassabis, D.: A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, *Science*, Vol. 362, No. 6419, pp. 1140–1144 (2018).

[21] Fujii, N., Sato, Y., Wakama, H., Kazai, K. and Katayose, H.: Evaluating human-like behaviors of video-game agents autonomously acquired with biological constraints, in *International Conference on Advances in Computer Entertainment Technology*, pp. 61–76 (2013).

[22] 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, Vol. 58, No. 3, pp. 167–173 (2013).

[23] 田中哲朗：「どうぶつしょうぎ」 の完全解析, 研究報告ゲーム情報学 (GI), Vol. 2009, No. 3, pp. 1–8 (2009).

[24] 中屋敷太一, 金子知適他：逆転の余地を考慮した評価関数の設計とどうぶつしょうぎによる評価, ゲームプログラミングワークショップ 2020 論文集, Vol. 2020, pp. 22–29 (2020).

[25] Song, Z. and Iida, H.: Using single conspiracy number for long term position evaluation, *ICGA Journal*, Vol. 40, No. 3, pp. 269–280 (2018).

[26] 将棋倶楽部 24 万局集:, ナイタイ出版 (2002).

[27] 齋藤雅史, 伊藤毅志：将棋 AI がプロ棋士の棋譜に与えた影響 ―定量的分析からの考察―, ゲームプログラミングワークショップ 2022 論文集, 第 2022 巻, pp. 159–166 (2022).

[28] 竹内聖悟, 林芳樹, 金子知適, 山口和紀, 川合慧：勝率に基づく評価関数の評価と最適化, 情報処理学会論文誌, Vol. 48, No. 11, pp. 3446–3454 (2007).

[29] Jacob, A. P., Wu, D. J., Farina, G., Lerer, A., Hu, H., Bakhtin, A., Andreas, J. and Brown, N.: Modeling strong and human-like gameplay with KL-regularized search, in *International Conference on Machine Learning*, pp. 9695–9728 (2022).

[30] Tsuruoka, Y., Yokoyama, D. and Chikayama, T.: Game-tree search algorithm based on realization probability, *ICGA Journal*, Vol. 25, No. 3, pp. 145–152 (2002).

[31] Obata, T., Sugiyama, T., Hoki, K. and Ito, T.: Consultation algorithm for Computer Shogi: Move decisions by majority, in *International Conference on Computers and Games*, pp. 156–165Springer (2010).