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)