氏名	小川竜欣
学位の種類	博士（工学）
学位記番号	共博工第 5 号
学位授与年月日	令和 8 年 3 月 25 日
論文題目	Data-Efficient Human-Like Policies at the Group and Individual Levels
論文審査委員	池田　心　北陸先端科学技術大学院大学　教授
	飯田　弘之　同　教授
	白井　清昭　同　教授
	米陀　佳祐　金沢大学　准教授
	鶴岡　慶雅　東京大学　教授

論文の内容の要旨

In this dissertation, we investigate methods for efficiently learning human-like policies under limited data availability. Here, a policy refers to a probability distribution over actions conditioned on states. We address both group-level and individual-level human-like behavior and also explore potential applications of such policies. One such application is human win rate prediction, which aims to estimate the probability that a human player would win the game if playing from a given position. We examine how human-like policies can be effectively applied in human win rate prediction. The primary target game is Shogi, a representative two-player perfect-information game for which large-scale game records are more difficult to obtain than in Chess. For broader comparison and validation, Go and Chess are also included as additional target domains.

First, we investigate the creation of human-like policies at the group-level perspective. We propose a Blend model that mixes the outputs of a supervised learning model and a reinforcement learning model so that human-like moves can be predicted even when human game records are limited. Experiments conducted on Chess, Go, and Shogi demonstrate that the proposed Maia-S, trained with roughly 1/100 of the 12 million-game dataset used in prior work (Maia), can achieve improved human move matching accuracies by combining the supervised model (Maia-S) with an AlphaZero-like reinforcement learning policy. Across the three games, the Blend model improves move-matching accuracy by 0.3–2.9 percentage points for intermediate players and 2.0–6.0 percentage points for advanced players. Additionally, in Chess, the Blend model improves move-matching accuracies by 0.2–1.1 percentage points even when using the original, large-data Maia model. In Go, the Blend model outperforms KL-regularized search, a method that improves move-matching accuracies by searching under calibrated supervised policies. While KL-regularized search itself improves upon Maia, the Blend model yields even greater gains. In Shogi, an analysis of positions where the Blend model is particularly effective reveals that, in positions where supervised models tend to suggest inferior moves, the Blend model more often aligns with human decisions. This indicates that blending compensates for the weaknesses of both supervised and reinforcement-learning-based policies, as originally expected.

Next, we examine how to realize human-like policies at the individual level. We address the challenge that many players have only a small number of game records available. To address this limitation, we propose the Similarity-Guided Fine-Tuning (SGFT) model, which leverages both the target player's games and games similar to them. Two methods for computing similarity are introduced: a feature-based approach using handcrafted game features, and an embedding-based approach using player-behavior embeddings. A two-stage fine-tuning framework is proposed: first fine-tuning with similar games, and then fine-tuning it with the target player's own games. In addition to move-matching accuracy – the primary metric used in prior studies – this dissertation introduces a new evaluation measure based on game features, designed to quantify how well a model imitates individual playing styles. Experiments on Shogi show that SGFT models outperform one-stage fine-tuning, which corresponds to the standard fine-tuning procedure where the target player's game records and similar game records are trained jointly. Moreover, embedding-based SGFT models outperform feature-based ones. The feature-based evaluation further reveals that even Transfer Maia, a prior

individual-level imitation method, does not significantly improve certain style-sensitive indicators. These results suggest that SGFT has strong potential for improving human-recognizable, style-specific characteristics.

Furthermore, we discuss the potential applications of human-like policies in playing, teaching, and commentary. A well-known issue in game AI is that AI-estimated win rates often diverge from the win rates that humans would actually achieve. Focusing on human win rate prediction, we propose an inner-product model. It computes the win rate by combining a policy with the predicted win rates of the successor positions. The proposed method surpasses existing models across multiple evaluation criteria – including result accuracy, cross-entropy, and expected calibration error – demonstrating the effectiveness of human-like policies for commentary applications in game AI.

In summary, we present an integrated framework for efficiently learning human-like policies under the practical constraint of limited data. The proposed methods are applicable across multiple games and provide a solid foundation for future research in human-like gameplay, educational AI, and commentary systems.

Keywords: human-like behavior, game AI, personalized modeling, supervised learning, reinforcement learning, fine-tuning

論文審査の結果の要旨

近年の AI 研究の進展に伴い，人間よりも強いコンピュータプレイヤを作るという目的がほぼ達成された今，本論文は，より人間らしいコンピュータプレイヤ，あるいはより“ある個人”らしいコンピュータプレイヤを作るための大きく分けて 3 つの内容からなる研究を行った．

1 つめは，特定の棋力帯のプレイヤらしい着手確率分布を求めるための研究である．既存手法 Maia は有力なアプローチであるが，学習対象ごとに 1200 万ゲームのデータを要しており，チェス以外のゲームではこのような大量のデータは期待できない．本論文では，データが少ない場合（具体的には 12 万ゲーム）に生じうる問題点を明らかにし，Maia による「人間的直感」と，AlphaZero 型学習による「読み」とを適切に組み合わせる Blend 推定法を提案した．この手法は，チェス，囲碁，将棋で評価され，データが少ない場合の着手予測正解率を既存手法よりも最大 3~7 ポイント向上させることが確認できた．本成果はゲーム関係のトップジャーナルである IEEE Transaction on Games に採録された．さらに，論文指導会の指摘を踏まえて，強さを模倣対象の人間プレイヤとほぼ等しくするための技術や，よりデータ数が小さい場合の提案手法の効果についても検証された．

2 つめは，特定の棋力帯のプレイヤについて，着手の確率分布ではなく，勝率の予測を正確に行うための研究である．囲碁や将棋のテレビ番組には現在の「勝率」が表示されることが多いが，この値はあくまで AI がその先を打ち継いだ場合のものであり，「人間にとっては勝ちにくい／勝ちやすい局面」があることが考慮されていない．本論文では，「人間が悪手を打ちそうな局面では勝率は下がるはず」といった仮説に基づき，着手確率分布と次局面の予測勝率の内積を取って新しい予測勝率とする手法を提案し，これが 0.5〜1 ポイントほど勝率予測の精度を向上させることを示した．本成果は国内最大のゲーム系会議 Game Programming Workshop に採録された．

3 つめは，プレイヤ群ではなく特定のプレイヤを模倣する研究である．Transfer Maia という既存研究では 5000 局対局したプレイヤの場合，Fine Tuning による精度向上（特化）が確認されている．本論文では，「もっと少ない対局数しかないプレイヤを模倣するために，そのプレイヤと似たプレイヤの棋譜を集めて追加学習すればよい」というアイデアをもとに，Embedding によるプレイヤ間類似性に基づいた追加学習と，2 段階ファインチューニングという手法の活用により，100 局しか対局していないプレイヤでも 1 ポイント程度の精度向上ができることを確認した．

以上から，博士（工学）の学位を与えるにふさわしいという判断で審査員が一致した．