審 査 の 結 果 の 要 旨

氏　　名　　水上　直紀

本論文は、「報酬が疎なゲームにおける強化学習を用いたプレイヤの構築」と題し、全 7 章から成る。強化学習は今日のゲーム AI 研究において最も重要なアプローチのひとつであり、多くのゲームでその有効性が報告されているが、報酬が頻繁に得られない場合には効果的な学習が非常に難しくなることが知られている。本論文は報酬のスパースネスの問題に対する対処法を複数提案し、ビデオゲームと麻雀においてその有効性を実験的に検証した論文である。

第 1 章は「はじめに」と題し、ゲーム AI 研究における強化学習の重要性、報酬が疎である場合に強化学習が難しくなること、および本論文において提案する手法の概略を述べている。

第 2 章は「報酬が疎な環境における UCB を用いた探索報酬の提案」と題し、深層強化学習において未知の局面を効果的に探索する戦略を提案している。具体的には、多腕バンディット問題に用いられる UCB 基準を探索のボーナスとして導入することで、方策勾配法の一種である A3C アルゴリズムの性能を向上させる手法を提案している。評価実験では、Atari と呼ばれるビデオゲームのシミュレータを用いた実験を行い、提案手法が既存の探索手法と比べて優れていることを報告している。学習の結果得られたエージェントは、強化学習にとって難しいゲームとされている Private Eye と Solaris において最高スコアを達成している。

第 3 章は、「牌譜からの教師付き学習によるコンピュータ麻雀プレイヤの構築」と題し、本論文が対象とする日本式麻雀のルールを説明し、牌譜を学習データとした教師付き学習によって四人麻雀のためのプレイヤを構築する手法を述べている。まず、上級者の牌譜データと平均化パーセプトロンアルゴリズムにより、単純にあがりを目指す一人麻雀プレイヤを構築し、それに、「降り」と「鳴き」の判断をするモデルを追加することで四人麻雀のためのプレイヤを実現している。

第 4 章は、「相手のモデル化とモンテカルロ法による麻雀プログラム」と題し、対戦相手の行動を予測するモデルとその活用法を説明している。提案手法では、

相手の状態を、「聴牌しているかどうか」、「待ち牌は何か」、「点数は何点か」の3つの観点でモデル化し、それぞれの観点について教師付き学習により予測モデルを構築している。さらに、それらの予測モデルを用いることで、四人麻雀を仮想的な一人麻雀に変換し、モンテカルロ法によって最善の手を決定する手法を述べている。

第5章は、「期待最終順位に基づくコンピュータ麻雀プレイヤの構築」と題し、一局単位の収支ではなく、一ゲームを通した最終的な順位を目的としたプレイヤを構築する手法を述べている。具体的には、多クラスロジスティック回帰モデルを用いて牌譜から教師付き学習を行うことで順位予測モデルを構築し、それをモンテカルロ法によるシミュレーションに導入することで、期待最終順位が最も良くなるであろう着手を行うプレイヤを実現している。インターネット麻雀サイトでの対戦実験の結果、提案手法によって点数状況の判断の精度が向上することが示されている。加えて、前章で説明したシミュレーションを改良するための手法を述べている。対戦相手の打点の予測モデルの改良、リーチが可能な局面での行動の改良、山に残っている牌の推定モデルの改良について述べている。

第6章は、「強化学習を用いた効率的な和了を行う麻雀プレイヤ」と題し、点数状況に応じて適切な役作りを行うための手法を提案している。最初に、強化学習によって手牌からあがれる翻数を予測するモデルを構築し、それを用いた探索によって最終的な順位の期待値が最善となる手を選択するという手法を提案している。

第7章は、「おわりに」と題し、本論文を総括し、今後の課題について述べている。

以上これを要するに、本論文はゲームAI研究において挑戦的な課題である報酬がスパースな状況での強化学習という問題に対して、未知局面に対する探索ボーナスを導入する手法、および最終的な報酬を予測するモデルを活用したエージェントの構築法を提案し、それらの有効性をビデオゲームと麻雀において実験的に明らかにしたものであり、情報処理工学に貢献するところが少なくない。

よって本論文は博士（工学）の学位請求論文として合格と認められる。