学位論文審査の要旨

博士の専攻分野の名称　博士 (情報科学)　氏名　高田　圭

審査担当者　主　査　教　授　山本 雅人
副　査　教　授　栗原 正仁
副　査　教　授　川村 秀憲
副　査　教　授　小野 哲雄
副　査　准教授　飯塚 博幸

学位論文題名

A Study on Learning Algorithms of Value and Policy Functions in Hex
（Hex を用いた局面評価関数とポリシー関数の学習アルゴリズムに関する研究）

ボードゲームを人間を超える強さでプレイするコンピュータプレイヤの開発は, グランドチャレンジの一つとして認識されてきた. 強いコンピュータプレイヤの開発を目的に, 先読みを含めた膨大な探索空間から最善手を効率よく探索する探索手法や, 自己対戦を通してコンピュータプレイヤを作成する機械学習アルゴリズムなどが提案されている. これらの手法は, 人工知能分野への貢献のみならず, 近年では材料科学や医学の分野への応用も期待されている.

本学位論文では,Piet Hein や John Nash らに開発された二人用ボードゲームである Hex を利用して大きく二つの機械学習手法を提案しており, それらの有効性をコンピュータプレイヤの開発を通して示している. 提案手法の一つは序盤や中盤といった局面状態を判別する分類器を Support Vector Machine (SVM) によって構築する手法である. 局面状態を分類することは, 局面状態に応じた探索によって良手を選択することが可能になることから, 多くのボードゲームで行われている戦略である. 既存手法では長い経験と知識に基づくルールで局面状態を分類するのが一般的だが, 著者らの提案手法ではより柔軟に局面状態を分類することを目指し,SVM で作成した分類器に基づいた局面状態の分類手法を提案した. 著者は, 作成した分類器に基づいて戦略を変更するコンピュータプレイヤを開発し, 既存のコンピュータプレイヤとの比較を通して, 構築した分類器が適切に局面状態を分類可能であることを示している. もう一つの提案手法は, コンピュータプレイヤの開発に必要な局面評価関数とポリシー関数を作成するための強化学習アルゴリズムである. 局面評価関数は, 局面の形勢を定量化する関数であり, ポリシー関数は候補手を評価する関数である. 提案手法と既存手法 (AlphaGo Zero や AlphaZero Algorithm) の主な違いはポリシー関数の学習方法にあり, 既存手法ではモンテカルロ木探索の探索結果である探索頻度分布を予測するようにポリシー関数を学習する. この手法は高精度なポリシー関数を作成可能ではあるが, 探索頻度分布を得るために各候補手に対する評価が多数必要であり, 学習に必要な計算コストが高い. 一方, 著者の提案手法では, 探索頻度分布ではなく, 局面評価関数による探索結果を予測するようにポリシー関数を訓練する. 探索頻度分布を利用する必要がないため計算コストの削減が期待できる. 提案手法で作成した二つの評価関数を使用するコンピュータプレイヤを開発し, 既存のコンピュータプレイヤとの比較から, 提案手法により非常に高精度な評価関数が作成可能であることを示している.

本学位論文は 5 章で構成されている. 第 1 章では, 研究背景および目的を述べ, 提案手法のアプ

ローチについて述べている. 第 2 章では, 研究に使用するボードゲーム Hex について,Hex のルール, 主な特徴, 局面のネットワーク化について, 代表的なコンピュータプレイヤ等の詳細を述べている. 第 3 章では, 複雑ネットワークの分野で得られた知見を利用し, 局面状態の分類器を構築した研究について述べている. まず,Hex の局面をネットワークとして捉え, ネットワーク特徴量を用いた大域的評価と局所的評価から構成される局面評価関数を提案し, その有効性を示している. 次に, 熟練者の棋譜を使用して SVM により局面状態の分類器を構築し, 局面状態に応じて戦略を変更するコンピュータプレイヤを開発している. 開発したコンピュータプレイヤと既存コンピュータプレイヤの比較から, 構築した分類器が適切に局面状態を分類可能であることを示している. 第 4 章では, 畳み込みニューラルネットワーク (CNN) で構成される局面評価関数とポリシー関数を作成するための強化学習アルゴリズムについて述べている. まず,Hex に CNN を適用する方法が述べられ,Hex における CNN の有効性を示している. 次に, 強化学習アルゴリズムを提案しており, 学習方法が異なる強化学習アルゴリズムと, 既存コンピュータプレイヤとの比較を通して, 提案手法の有効性を明らかにしている. 第 5 章では, 本学位論文の結論を述べている.

これを要するに, 著者はコンピュータプレイヤの開発に必要な評価関数を作成するための学習アルゴリズムを提案し, その有効性を示した. 特に, 著者の提案した強化学習アルゴリズムは, 非常に高精度な評価関数を作成可能であり, 他のゲームにも適用可能な汎用的なアルゴリズムである. これらの成果は, 機械学習やコンピュータプレイヤの開発に対し有益な知見を得たものであり, 人工知能分野における研究に貢献するところ大なるものがある. よって著者は, 北海道大学博士 (情報科学) の学位を授与される資格あるものと認める.