学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称　博士（情報科学）　氏名　高田　圭

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）

ボードゲームをプレイするコンピュータプレイヤの開発は, エンタテイメント性の追求だけではなく, 現在の局面から次手を決定する探索問題への研究でもある. 先読みを含めた膨大な探索空間から, 自分の目的関数を最大化する最善手を探索する手法は, 人工知能分野や最適化問題などに応用される探索手法の確立に貢献してきた. 本研究の目的は, 優れた手を探索するために必要となる 2 つの評価関数を作成する機械学習アルゴリズムを開発することである. 必要となる 1 つの評価関数は, 局面の形勢を定量化する局面評価関数であり, もう 1 つは候補手の有望性を定量化するポリシー関数である. 本研究では,Piet Hein や John Nash によって開発された二人用ボードゲーム Hex を利用し, 局面状態の分類器を機械学習から作成する手法と, 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を提案し, それらの有効性を明らかにする.

序盤や終盤といった局面状態の分類は, 局面状態に応じてコンピュータプレイヤの戦略を動的に変更することが可能になるため, 多くのゲームで行われている. 局面状態の分類手法として, 将棋では駒同士がぶつかるまでを序盤, それ以降中盤などの事前に定めたルールに基づいて局面状態を分類する手法がある. しかし, これらの手法は例外的な局面に対応することが困難である欠点をもつ. 本研究では, より柔軟に局面状態を分類し, その分類結果に基づいて戦略を適切に変更することが可能なコンピュータプレイヤを開発することを目的に,Support Vector Machine (SVM) を用いた局面状態の分類器を開発する. まず, 戦略の変更が可能なプレイヤを開発するために, 局面評価関数を提案する.Hex の局面はセルをノード, セルの隣接関係をリンクとすることで, ボードネットワークとして捉えることが可能である. 既存局面評価関数は, ボードネットワークを利用して局面を 1 つの観点から評価してきた. 提案局面評価関数は, 複雑ネットワークの分野で提案されているネットワーク指標を用いることで,2 つの観点から局面を評価する. 提案局面評価関数を使用するコンピュータプレイヤの戦略は,2 つの評価指標の評価比率により決定される. 戦略を変更するべき適切なタイミングを決定するために, 熟練者の棋譜を使用して, 序盤と中盤以降を分類可能な分類器を SVM で作成する. 作成した分類器が適切に局面状態を分類可能であることを明らかにするため, 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, 代表的な他のプレイヤとの比較実験を行った. 実験結果から,SVM で作成した分類器は局面状態を適切に分類可能であることを示した.

ここまでの研究では熟練者の棋譜を利用した手法を提案してきたが, 一般的に熟練者の棋譜を用いた学習手法では, 熟練者を超えるコンピュータプレイヤを開発することは困難である. 熟練者を超えるコンピュータプレイヤの開発を目的に, 局面評価関数とポリシー関数を作成する強化学習アルゴリズムが提案されている.Silver らは, 自己対戦を繰り返す強化学習アルゴリズム (AlphaGo Zero, AlphaZero Algorithm) を提案し, 囲碁において人間のトッププロ棋士を打ち破るコンピュータプレイヤを開発した. この手法では, 局面評価関数は勝敗を予測するように訓練され, ポリシー関数はモンテカルロ木探索による各候補手の探索頻度を予測するように訓練される. この手法は非常に高精

度な局面評価関数とポリシー関数を作成可能であるが, 一方で, 候補手の探索頻度分布を得るためには多数回のシミュレーションが必要であり, 学習に必要な計算コストは高い. 本研究では, 探索頻度分布を必要としない強化学習アルゴリズムを提案する. 提案アルゴリズムでは, 勝敗結果と探索結果の指し手を使用して 2 つの評価関数を訓練するため, 必要なシミュレーション回数と計算コストの削減が可能となる.2 つの評価関数は, 畳み込みニューラルネットワーク (CNN) で構成され, 局面評価関数は勝敗結果を予測するように訓練され, ポリシー関数は良手を探索対象とするように訓練される. 提案アルゴリズムにより高精度な局面評価関数とポリシー関数が作成可能であることを示すために, 訓練された 2 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実験を行った. 提案したコンピュータプレイヤは,2017 年世界 1 位のコンピュータプレイヤに対して, 同じ探索条件下で約 80% の勝率を示し, 提案アルゴリズムは非常に高精度な局面評価関数とポリシー関数を作成可能であることが示された.

本論文は, 全 5 章で構成される.1 章では, 研究背景及び目的を述べる.2 章では, 本研究で使用するボードゲーム Hex について述べる.Hex の局面のネットワーク化手法, 代表的なゲーム木探索アルゴリズム,Hex における主なコンピュータプレイヤと,Hex 特有の手法等についてまとめている.3 章では,Hex に対して複雑ネットワーク理論を応用し, 局面状態の分類器を作成した研究について述べている.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する. 既存のコンピュータプレイヤとの比較を通して, 大域的・局所的評価を組み合わせることが有効であることを明らかにする. また,SVM を使用し, 局面状態を判別する分類器を作成する. 作成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すために, 提案手法を使用するコンピュータプレイヤと既存のコンピュータプレイヤとの比較実験を行う.4 章では,CNN を使用した局面評価関数とポリシー関数を開発し, 自己対戦を通して 2 つの関数を訓練する強化学習アルゴリズムを提案している.CNN によるポリシー関数とネットワーク特徴量によるポリシー関数の比較を通して,Hex においても CNN が有効であることを明らかにする.
また, 提案する強化学習アルゴリズムについて述べ, 提案アルゴリズムで訓練された 2 つの関数が高精度であることを, 既存のコンピュータプレイヤとの比較から明らかにする.5 章では, 本論文の結論を述べる.