

氏 名 若原 拓己
学 位 名 博士（システム情報科学）
学位記番号 第21号
学位授与年月日 平成25年3月21日
学位論文題目 長遅延報酬対象向け強化学習手法

論文審査委員 主査 三上 貞芳
副査 大澤 英一
副査 片桐 恭弘
副査 松原 仁

論 文 要 旨

強化学習手法は外乱に強く、教師データが不要な機械学習手法の一つである。そのため強化学習手法は不確実性が高い実用的問題に適していると考えられる。実用的問題において強化学習手法を適用させることを考えると、制御に対し報酬は大きな時間遅れを持って与えられ、また制御の回数と報酬を得る回数にも差が生じることが考えられる。従来の強化学習手法では状態行動対と報酬は厳密に関係を結びつけることが可能である対象に適用するような設計が主なため、従来の強化学習手法をそのまま実用的な問題に適用させるのは難しい。そのため本研究では実用的問題に向けた強化学習手法の提案を行った。実用的問題の制御に対し報酬が大きな時間遅れを持って与えられるという特徴から、大まかな状態行動対を報酬が得られるまでの期間とし、その状態行動対は実際に行われる制御の集合とした。提案手法は後述する植物育成制御問題に強化学習手法を適用させることを基盤としており、状態行動対の集合と報酬の関係は植物育成過程の特徴を基に考案した。提案手法は状態行動対を集合として扱い、報酬は得られた時点で一括してその集合に与えられる。そのため報酬を得られた時点でそれまでの状態行動対は一括して強化されることになり、それにより学習速度が速まることが植物育成制御問題への適用で検証済みである。

植物育成制御問題と交通流制御問題を実用的問題の例として提案手法を適用させる。

植物育成制御問題について、工学的に植物育成制御を行うことが可能な植物工場システムがある。植物工場システムは大別して人工環境だけでまかなう完全制御型、一部自然環境を用いる半制御型に分けられる。植物の育成モデルは確立されていないため、自動制御により植物育成を行うのは難しい。本研究では提案した学習手法を用いて植物工場システ

ム下で植物育成制御を行う。植物育成制御は異なる成分比を持った養液の供給比率制御により行う。実験環境としてインキュベータを用いた小規模植物工場システムを作成し、従来の強化学習手法による育成実験を行った。この実験では従来手法による制御では育成回数を重ねることで改善される結果は得られなかった。提案手法と従来手法の比較実験は植物育成シミュレータを作成し行った。シミュレータ実験では提案手法が従来手法に比べ効果が高いことを確認した。

交通制御問題について、都市交通の渋滞解消のための交通信号制御方式として、交通流の流入量と流出量の期間を離散化した交通ネットワークモデルを扱うストアアンドフォワード方式に基づいた手法がある。交通流を離散化することにより、各道路の流入流出日による交通信号のスプリットのフィードバック制御を適用させることが可能であり、LQ最適化手法といった最適制御を用いることでフィードバックゲインを導き出すことが可能である。この方法は理論的・実用的観点から有用であり、いくつかの都市で使用されているがフィードバックゲインを導出するためのパラメータは人の手で事前に決定されている。そのパラメータの一つに分岐率がある。分岐率とは直進右左折と言った交差点での自動車の進行方向の割合を表すもので、運転手の意思という人的要素が絡みセンサで計測することは不可能であるが、フィードバック制御を行うことを考えた場合大きな影響を及ぼす要素となる。本研究では分岐率の推定を強化学習手法で行い、その予測値を使用した交通信号のフィードバック制御による渋滞解消を目指す。検証は小規模な交通ネットワークのミクロシミュレータを作成し、それを用いて行った。シミュレータでは2交差点の交通ネットワークにおいて、片側交差点の信号のみをフィードバック制御を行いもう一方は固定制御で行った。フィードバック制御の有無で比較し、強化学習により分岐率を獲得することとそれを用いたフィードバック制御の有効性を確認した。

審査結果の要旨

本論文は、知能システムの分野の研究テーマの一つである、数理モデルに基づく制御が困難であるシステムに対して、経験を通じて行動方策を獲得させることで目標に近づける手法の一つである、強化学習手法の研究に取り組んでいる。強化学習によるシステムの制御については、従来から仮想世界やシミュレーションを対象とした理論面の研究が多く行われているが、実用的な時間で有用な行動方策を得ることが望まれる実応用については、まだ十分な研究が進んでいない。

特に、行動方策を獲得させるためのシステムの改善度合いを評価する値（報酬値）が、行動の後大きく時間遅れで得られるような問題は、数理モデル化が困難な実応用で多く見られるが、従来から強化学習の適用で良好な行動方策を実用的な時間で得ることが難しい対象であった。

本論文では、多くの実問題で「イベント」と本論で名づけた特徴的な状態行動が見出せることを示し、これにより上記の問題を解決する手法を提案している。「イベント」とは、設計者が容易に見いだせるような、システムの振る舞いが大きく変わる状態行動であり、多くの問題では、隣接するイベントで囲まれた状態遷移が後戻りを含まないように設計できることが、本論で示されている。この前提が成り立つ場合には、イベントで囲まれた状態遷移を一つの状態行動のように集合として扱うことで、その内部を強く強化し、全体を従来型の強化学習で強化することで、誤った行動を強化することなく、実問題で望まれるような有用な方策を早い時点で得ることができることを示している。

続いて、代表的な問題として植物工場での育成制御問題について取り上げ、養液の混合比の制御による学習育成手法を述べ、実際の小規模閉鎖型植物工場での実験、およびシミュレーション実験での結果を示して本方法の有用性と適用可能性を明らかにしている。

さらにもう一つの代表的な問題として、交通信号機の制御問題を取り上げている。状態フィードバックによる区域制御において、重要かつ測定困難な対象である交差点での分岐率を推測する問題に、提案手法を適用する例を示し、交通流シミュレーションによる実験で有用性を明らかにしている。

以上のような研究の展開、成果は、知能システム分野で従来取扱いが困難であった対象を扱う新規手法の提案、および有用な実事例での適用方法を明らかにするもので、システム情報科学の発展に貢献するものである。よって本論文は博士（システム情報科学）の学位授与に値するものと判断する。