

氏 名（国籍）	アグス ナバ（インドネシア）		
学 位 の 種 類	博 士（工 学）		
学 位 記 番 号	博 甲 第 4279 号		
学位授与年月日	平成 19 年 3 月 23 日		
学位授与の要件	学位規則第 4 条第 1 項該当		
審 査 研 究 科	システム情報工学研究科		
学位論文題目	Adaptive Control based on Adjustable Evaluation Function (調整可能な評価関数を用いた適応制御に関する研究)		
主 査	筑波大学教授	工学博士	西 原 清 一
副 査	筑波大学教授	工学博士	宮 本 定 明
副 査	筑波大学教授	工学博士	安 信 誠 二
副 査	筑波大学教授	博士（工学）	福 井 幸 男
副 査	筑波大学助教授（連携大学院） (産業技術総合研究所)	工学博士	宮 下 和 雄

論 文 の 内 容 の 要 旨

制御設計は、制御対象であるプラントが望みどおりに挙動するようにコントローラを操作することを主要な課題としている。適切に定義された問題に対しては多くの有効な手法が提案されてきた。しかし、非線形性を含むより広い問題に対して従来手法は適用できないことも多い。これには、適応制御の手法が広く用いられてきた。

多くの場合、プラントの出力エラー、すなわち、プラントの状態と目標との間の差を最小化することが適応制御の目標である。しかし、必ずしもその指標が常に直接的に正しい結果を与えとは限らず、ときには誤ったゴール（非目標状態）へ導いてしまうこともある。出力誤差の小さいことが即、評価情報の正当性を意味しないからである。

強化学習（RL）では、行為（アクション）の評価情報は必ずしも既知あるいは直ちには得られないものという前提がある。RL では行為の評価値は、技術者（教師）によって事前にではなく、環境によって後刻‘報酬’として与えられる。この報酬は、良さ（goodness）を表す価値関数で規定される。この価値関数は、行為を繰り返しつつ報酬とペナルティの経験を積むという時間をかけた学習によってその精度を増していく。RL の主要な課題は、この試行錯誤を時間をかけて行なう必要があるという点である。

本論文では、著者は制御設計のための代替的なアプローチを示している。提案手法は、ポリシーすなわち、状態から行為へのマッピングを、制御変数とは独立の関数で表現するという強化学習の一手法を取り入れている。このポリシー探索手法は、学習される価値関数によって測られる制御能力が改善されるようにポリシーパラメータを調整する手法である。

提案されたアプローチは、近似評価関数で測られるコントロール性能を、オンラインで調整しながら向上させるという方法である。このアプローチにより、評価関数の調整には長時間かけてアクション実施結果が得られるのを待つ必要はなく、状態、アクション、アクション実施結果などの容易に得られる情報を直接的に用いることができる。

以上、本論文では、ポリシー探索法を台車付き倒立振子の問題に適用し、シミュレーション実験を行ない本アプローチの有効性を示している。評価関数の近似値を与え、また、近似誤差をオンラインでチューニングすることができるという点で、実用的かつ効果的な設計への可能性を示したものである。

審 査 の 結 果 の 要 旨

制御設計の問題に対して、知識処理における強化学習の手法を取り入れて、実際の台車付き倒立振子へ適用し、シミュレーション実験を実施した。この一貫した研究の推進を通して、強化学習の一つであるポリシー探索手法の有効性について実証を試みた点が評価できる。今後は、実システムにおける検証および、本手法の信頼性や効率についてより精確な評価が課題として残されている。

よって、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。