

氏名(本籍)	小林 高彰		
学位の種類	博士(工学)		
学位記番号	博甲第	7293	号
学位授与年月日	平成 27 年 3 月 25 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	選択的不感化ニューラルネットを用いた 連続状態行動空間における強化学習		
主査	筑波大学 教授	工学博士	森田 昌彦
副査	筑波大学 教授	工学博士	安信 誠二
副査	筑波大学 准教授	博士(工学)	掛谷 英紀
副査	筑波大学 准教授	博士(工学)	田中 文英
副査	筑波大学 助教	博士(工学)	澁谷 長史

論文の要旨

制御対象を十分にモデル化できない問題に対して、システムが対象に自律的に適応して制御を実現する方法として強化学習がある。よく用いられる強化学習のアルゴリズムはQ学習などのCritic-Onlyの手法とActor-Critic手法に大別できるが、実問題のように状態・行動がともに連続な問題には、Actor-Critic手法を適用するのが一般的であった。しかし、Actor-Critic手法はアルゴリズムが複雑で学習が遅いなどの問題がある。一方、Critic-Onlyの手法は、あまりよい関数近似手法がない上に、価値が最大となる最適行動を求めるのに多大な計算コストを要するため、これまで連続な状態行動空間にCritic-Onlyの手法を適用する実用性の高い方法はなかった。

本論文では、選択的不感化ニューラルネット(SDNN)を行動価値関数の近似器として用い、多数の行動に関するQ値を同時に出力層に分散表現することによって、連続な状態行動空間にCritic-Onlyの手法を適用する実用的な方法を提案する(第2章)。

この手法を、制約のあるアクロバットの振り上げ制御課題に適用し、既存の関数近似器を用いた場合と比較した。その結果、提案手法によって計算コストの増加が大幅に抑えられ、実時間制御に使えること、離散行動の場合よりも環境の連続的な変化に追従しやすいことなどが示した(第3章)。また、冗長な状態変数を含む場合について実験を行った結果、提案手法を用いると、冗長変数の影響をほとんど受けないことを明らかにした(第4章)。さらに、制御周期を短くすることができず、行動値を連続にしないと制御困難な場合についても実験を行い、提案手法がこのような状況下でも有効であることを確認した(第5章)。最後に、これらの結果に基づいて、状態行動空間が連続な場合のQ学習に関する主要な先行研究との比較を行い、提案手法の特長等について考察した(第6章)。

結論として、本提案手法は、計算コストの増加や学習効率の低下といった既存のCritic-Only手法

の実用上の問題点を解決することができ、冗長な状態変数に対してロバスト性を有するとともに、従来手法では解くことができない条件下でも適用可能である。また、アルゴリズムが単純でパラメータ設定も容易であるといった点でも実用性が高い。以上のことから、本手法は、強化学習の適用範囲を大きく広げるものと言える。

審 査 の 要 旨

【批評】

連続な状態空間において Q 学習を行う場合、価値関数の近似が不可欠であるが、強化学習において通常用いられる局所的な関数近似手法では、近似精度と学習効率とを両立させることができない。行動空間も連続な場合には、最適行動 (Q 値の最大値) を求めるのに多大な計算コストを要するため、これまで実用性の高い方法はなかった。本研究では、選択的不感化ニューラルネット (SDNN) を行動価値関数の近似器として用い、多数の行動に関する Q 値を同時に出力層に分散表現することによってこの問題の解決する方法を提案している。アクロボットの振り上げ課題について実験を行った結果、既存の関数近似器を用いた場合に比べて計算コストの増加が大幅に抑えられ、実時間制御に使えることが明らかになった。また、離散行動の場合よりも環境の連続的な変化にも追従しやすいこと、冗長変数に対する高いロバスト性をもつこと、従来手法では解くことができないような問題でも解ける場合があることなどが示された。

このように、本論文は、状態だけでなく行動も連続である問題に適用可能な Q 学習の手法を提案し、その有効性を検証したものである。提案された手法は、問題によっては既存の手法よりはるかに高性能を示すだけでなく、アルゴリズムが単純でパラメータ設定も容易であるなど実用性が高い。従って、強化学習を適用できる問題の領域を大きく拡大するものであり、その点において特に高く評価できる。今後、より多くの課題に適用することによって、工学的な有用性を確立するとともに、現実問題へ応用することが期待される。

なお、本論文の第 4 章以外の主要な内容は、厳正な査読を受けた上で電子情報通信学会論文誌に採録されていること、また第 4 章の内容は査読付き国際会議で発表済みであることを確認した。

【最終試験の結果】

平成 27 年 1 月 27 日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

【結論】

上記の学位論文審査ならびに最終試験の結果に基づき、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。