

論文内容の要旨

博士論文題目 パターン生成器を用いた制御に対する強化学習法

氏名 中村 泰

(論文内容の要旨)

実空間で行動するロボットに対する制御を行う場合、環境の知識が前もって得られなかったり地面などの環境が動的に変化するため、あらかじめ与えられた制御を行うのではなく自律的に環境に適応し、制御則を獲得する枠組みが必要となる。強化学習法はこの要求を満たす手法で、環境や制御対象のシステムなどの先見的な知識が無い場合にも適用できる。しかしながら、歩行などの複雑な運動に対する制御則の学習は、システムの自由度が大きく非線形性が非常に強いことや、制御則自体が非線形で複雑となることから困難なものとなる。

一方、歩行などの生物の運動は、周期的な信号を生成する中枢パターン生成器 (CPG) とよばれる神経回路によって制御されていることが示唆されている。このような生物の制御機構を参考にして、周期的な運動に対する CPG を用いた制御法の研究が行われてきた。これらの手法では、制御器となる CPG を再帰結合型ニューラルネットワークで実装し、制御対象のシステムからのフィードバック結合やニューロン間の相互結合の重みを変化させることによって制御信号を調整する。

本論文では、CPG を用いた制御に対する自律的な学習の枠組みとして、CPG-actor-critic モデルと呼ばれる新しい強化学習法を提案する。提案手法は、複雑な制御則を学習する代わりに CPG の結合重みの学習を行う手法であり、線形制御器に対する学習を行うことで非線形な制御則を獲得できる。

まず、価値関数に基づいて学習を行う CPG-actor-critic モデルの学習法を導出する。提案手法を 2 足歩行ロボットの自律的な歩行運動の獲得課題に対する計算機シミュレーションに適用し、提案手法により安定した 2 足歩行を実現する CPG コントローラを獲得できることを示す。また、学習によって得られた CPG コントローラの性質について報告する。

次に、方策勾配法に基づいて学習を行う CPG-actor-critic モデルの学習法を導出する。近年、学習の収束が保証された方策勾配法を用いた actor-critic モデルが提案された。価値関数に基づく手法ではシステムの複雑さによる学習の困難さは克服できていないが、方策勾配法を用いた手法ではこの困難が克服できる。提案手法を 2 足歩行ロボットの自律的な歩行運動の獲得課題に対する計算機シミュレーションに適用し、提案手法により安定した 2 足歩行を実現する CPG コントローラを獲得できることを示し、またパラメータの改善に対して良い性質を持つことを示す。

(論文審査結果の要旨)

強化学習法は試行錯誤を行うことにより自律的に制御則を獲得する枠組みで、理論的な研究や工学的な応用が行われてきている。また、生物の運動は環境の変化に対して柔軟に適応できるという特徴を持っており、生物の運動制御機構をモデル化した CPG を用いた制御法も工学的に研究されている。本論文は、設計が困難な CPG のパラメータを自律的に決定する学習手法を提案したもので、2 足歩行ロボットシミュレータに対する制御則の獲得問題に適用し、提案手法の有効性を示した。本論文の主な成果は以下のように要約される。

1. CPG コントローラを分割することにより、再帰結合型ニューラルネットワークの学習と部分観測問題という困難を回避した新たな強化学習法の枠組み (CPG-actor-critic モデル) を提案した。
2. CPG-actor-critic モデルに対する価値関数に基づいた強化学習法を導出した。また、提案手法を 2 足歩行ロボットシミュレータの歩行運動獲得問題に適用することにより、学習によって安定した歩行が可能な制御則が獲得できることを示した。
3. 価値関数に基づく強化学習法において課題となった学習過程の不安定さを改善するため、方策勾配法に基づいた CPG-actor-critic モデルを導出し、2 足歩行ロボットシミュレータの歩行運動に対する制御則が安定した学習過程により学習可能であることを示した。さらに、方策勾配法の学習速度の向上のために、自然方策勾配法に基づいた学習法を定式化し、2 足歩行ロボットシミュレータの歩行運動獲得問題により、その有効性を示した。また、新しい環境で学習することにより、環境に適応できることを示した。

以上のように、本論文では生物のリズム運動制御機構を参考にした制御法に対して自律的な学習を可能とする強化学習法を提案し、2 足歩行シミュレータを用いた歩行運動獲得課題により有効性を示した。また、方策勾配法を用いることにより自由度の大きなシステムである 2 足歩行ロボットシミュレータに対しても安定した学習が行うことができ、新しい環境にも適応できることを示した。

本論文の成果は、自由度の大きなシステムに対する自律的な制御則獲得のための有効な知見を与えるものであり、重要な研究であると評価できる。工学、特に適応制御の分野において学術上、および、応用上の寄与が十分であるため、博士(工学)の学位論文として価値のあるものと認める。