

論文内容の要旨

博士論文題目 統計的手法による関数近似法と強化学習への応用に関する研究

氏名 吉本 潤一郎

(論文内容の要旨)

人間の情報処理過程を考える上で、その最も重要な性質は未知の環境でも経験に基づいて学習・適応する能力を有することである。このような情報処理を計算機により実現しようとするのが機械学習である。機械学習は、問題の定式化により教師あり学習、教師なし学習、強化学習に分類することができる。教師あり学習は、入力データと各入力に対応する理想的な出力データからその入出力関係を良く再現するための学習であり、関数近似問題が対応する。教師なし学習は、入力データのみからその特性を自己組織的に獲得する学習であり、主成分分析法や因子分析法などの特徴抽出問題やラベルなしデータからのクラスタリング問題が対応する。強化学習は、試行錯誤を通して報酬を最大化する制御を獲得する問題であり、最適制御問題が対応する。いずれの学習スキームも工学における重要な問題を含んでおり、その研究意義は大きい。

これらの学習スキームは、データの生成過程を確率分布として記述することにより、統計的推定問題として統合して扱うことができる。本論文では、統計的推定問題の観点から、関数近似問題と最適制御問題への応用を目的とした新たな機械学習法を提案する。

本論文では、関数近似モデルとして正規化ガウス関数ネットワーク (Normalized Gaussian network, NGnet) が用いられる。NGnet は、正規化されたガウス関数を用いて入力空間を滑らかに分割し、分割された部分空間ごとに線形回帰を行うモデルである。NGnet は入出力変数の同時分布に対して確率モデルを定義することにより、統計的推定法の一つである Expectation-Maximization (EM) アルゴリズムを用いて効率良く学習することができる。一方で、EM アルゴリズムはバッチ学習法であるために、実時間でのモデル同定問題や入出力分布が時間とともに変化する動的な関数近似問題への応用が困難である。この点を改善するために、オンライン学習が可能な EM アルゴリズムが導出される。

次に、オンライン EM アルゴリズムを用いた連続力学システムの自動制御のための強化学習法が提案される。強化学習では、試行錯誤を通して各状態および行動に対する良さを評価するための価値関数が学習され、それに基づいて方策と呼ばれる制御関数の学習が行われる。これらの2つの関数は相互依存の関係にあり、並行して学習が行われるために動的な関数近似問題と対応している。したがって、オンライン EM アルゴリズムが有効に働くことを期待できる。提案された強化学習法は、倒立振り子と acrobot の自動制御問題に応用され、その性能が評価される。

NGnet のような非線形混合モデルの学習において、最適な混合数をどのように決定するか、すなわち、モデル構造の決定は重要な問題の一つである。この点を解決する手法として、ベイズ推定法に基づく学習法が提案される。また、効率良くモデル構造選択を行うための階層的モデル選択手法が提案される。提案された学習法は、非線形力学システムの同定問題や強化学習における関数近似問題に応用され、その有効性が示される。

(論文審査結果の要旨)

与えられた問題を学習によって解決するアプローチは、性質が未知な環境をモデル化する上で、また、人間の情報処理過程をモデル化する上で重要な工学的問題である。本論文は、統計的推定問題の観点から正規化ガウス関数ネットワーク (NGnet) による関数近似法を提案し、強化学習における価値関数や最適方策の近似問題に応用したものである。本論文の主な成果は以下のように要約される。

1. EM アルゴリズムに基づく NGnet のオンライン関数近似法を強化学習における価値関数の学習に適用することにより、連続力学システムのための新たな強化学習法を提案している。また、提案した強化学習法を困難な制御課題である acrobot の自動制御問題に適用することにより、その有効性を示している。
2. 強化学習法において学習結果に基づく良い行動を実現するための「貪欲な行動」と様々な空間を試してみるための「探索的な行動」のトレードオフをいかに解決するかは効率良い学習を実現する上で重要な問題である。その解決方法として「ボルツマン行動選択則」を取り上げ、NGnet を用いて効率良く近似する手法を最尤推定法に基づいて導出し、計算機シミュレーションによりその有効性を示している。
3. 複雑な関数近似モデルを用いることによる過学習を避けるために、ベイズ推定法に基づく NGnet の学習法と最適モデル構造を選択するための評価基準を導出している。さらに、効率良いモデル構造選択を行うための階層的モデル選択手法を提案している。また、関数近似問題と非線形力学システムの同定問題に提案手法を適用することによってその有効性を示している。

本論文の成果は、未知の環境下、あるいは、時間とともに変化する動的な環境下における適応システムや自動制御のための重要な知見を与えるものであり、工学、特に、適応制御と学習理論の分野において学術上、および、応用上の寄与が少なくない。よって、博士(工学)の学位論文として価値のあるものと認める。