

論文内容の要旨

博士論文題目

モジュール構造を利用したニューロコントローラの進化的構造探索法および複数報酬強化学習法の提案

氏名

上岡 拓未

自律システムが未知環境に対応するためには、機械学習や進化的計算法といった最適化手法によって、環境に応じた制御器を獲得することが必要である。このような最適化手法を具体的問題へ適用する際には手法の選択だけでなく、制御器の構造、目的関数、メタパラメータなどの設計事項を決定する必要がある。多くの場合、こうした設計は実験者の直感と試行錯誤によって行われ一方で、ロボットの制御などの複雑な問題では特に制御器の構造と目的関数の設計は重要な問題となる。そこで、本研究では、(1) 制御器の構造の自律的な獲得手法と、(2) 目的関数の設定に対して頑健な学習手法を提案する。制御器の構造を自律的に獲得する手法として、ニューラルネットワークの進化的探索法である NeuroEvolution (NE) と呼ばれる枠組みがある。これまでに提案されてきた多くの NE 手法は 1 つの遺伝子に対して 1 つの結合重みを割り当てていたため、結合単位の探索しか行うことができなかった。そのため、大規模なネットワークへの適用は困難であった。本研究ではモジュール単位での構造探索を可能にすることで、大規模なネットワークにおける探索効率を改善した。提案した遺伝的表現ではニューロン間の結合に実数値の重みとネットワークの部分構造のどちらでも区別なく割り当てることが可能である。これにより、ネットワークをモジュールの組み合わせで表現でき、結合単位の探索とモジュール単位の探索が同列に扱える。シミュレーション実験により、提案手法がモジュール構造を利用した構造探索によって非マルコフ問題を解けることを示した。また、強化学習を多目的最適化問題に適用する際には、各報酬値の組み合わせを考慮する必要があるため、報酬設計が困難になるという問題があった。そこで、本研究では報酬値の組み合わせに対してロバストな強化学習法 Max-Min Actor-Critic (MMAC) を提案した。MMAC は目的別に与えられた報酬関数ごとに価値関数を学習する同一構造のモジュールを持ち、各状態における最小の価値関数を最大化する方策である Max-Min 最適方策を得る。シミュレーション実験によって、Max-Min 最適方策は従来の強化学習法に比べて各報酬値の組み合わせの影響を受けにくいことを示した。これにより、多目的最適化問題における報酬関数の設計問題を改善した。

(論文審査結果の要旨)

機械学習や進化的計算による制御器の最適化には様々な応用例が報告されている一方で、制御器の構造や目的関数などは実験者の直感と試行錯誤によって決められる場合が多い。設計事項の試行錯誤によらない解決法は、真に自律的なシステムの実現に向けて不可欠である。本論文では制御器の構造と目的関数の設計問題を軽減することを目指している。本論文の主な成果は以下のように要約される。

1. モジュール構造を利用した構造探索を可能にしたニューラルネットワークの遺伝的表現方法を提案した。本手法はネットワークの重み結合と部分構造を同様に扱うことにより、従来手法よりも短い遺伝子長によって大規模なネットワークの表現が可能である。シミュレーション実験により、本手法がモジュール構造を利用した構造探索によって非マルコフ問題に対して有効であり、また従来手法では困難であった大規模な制御器の構造進化が可能であることを示した。
2. 複数報酬における強化学習手法である Max-Min Actor-Critic 法を新規に提案した。本手法は、複数の報酬に対してそれぞれ別のモジュールによって価値関数を学習し、行動器において Max-Min 方策を学習する。シミュレーション実験により Max-Min 方策は報酬値の組み合わせの影響を受けにくいという意味でロバストであることを示した。これにより、強化学習を多目的最適化問題に適用する際の報酬関数の設計問題が改善された。

以上のように本論文では、最適化手法を実問題に応用する際の設計問題を改善することで、設計者に依存しない真に自律的なシステムの実現に大きく貢献している。よって、博士（工学）の学位論文として価値あるものと認める。