

## 審査の結果の要旨

氏 名 森 口 博 貴

本論文は、「Policy Search and Model Learning in Reinforcement Learning via Evolutionary Computation（進化計算を用いた強化学習における政策探索とモデル学習）」と題するもので、ロボットにおける強化学習の諸課題の解決を目的として、効率的かつロバストな学習を可能とする技術について論じたものである。

ロボットにおける強化学習では、連続的な行動・状態空間を対象とするため、古典的な離散的状態・行動空間を対象とする強化学習手法を単純に応用することは適切でない。これに対して本論文では、進化計算の柔軟性に着目し、高次元な状態・行動空間を持つ問題に対しては進化型ニューラルネットを用いた政策探索の効率性・ロバスト性を改善させる一方で、低次元な状態・行動空間を持つ場合にはシンボリック回帰を用いたモデル学習で超高サンプル効率を実現することで、様々な種類のロボット応用に応じた一連の強化学習アルゴリズムを提案している。

各章の具体的な内容は以下のとおりである。

第1章「Robot Automation and Reinforcement Learning」では、ロボットの自律制御への応用を志向した強化学習において、古典的な離散的状態・行動構造を想定する場合とは異なる問題が生じることを概説している。そして、本論文で特に解決すべき課題として、(1) 連続値かつ高次元の状態・行動空間の扱い、(2) 動力学モデルが不明である場合の扱い、(3) サンプル効率の向上、(4) 局所最適解への収束に対するロバスト性の向上、の4点を掲げている。

第2章「Policy Search and Model Learning」では、第1章であげた強化学習の問題点を解決すべく、これまで提案されてきた政策探索とモデル学習の手法について概説・比較している。その中でも特に本論文が対象とする、進化型ニューラルネットを用いた政策探索とシンボリック回帰を用いたモデル学習について、既存手法における課題とそれらの解決方針を議論している。

第3章「Robust Neuroevolution through Sustaining Behavioral Diversity」では、ニューラルネットの環境下での振る舞いに基づくニッチ化を用いることで、進化型ニューラルネットの一種であるNEATのロバスト性を高める手法を提案している。この手法ではこれまでに実現

例の無かったNEATでの振る舞い情報の利用を可能としており、また類似研究に比べて詳細な実証実験によって、提案手法の有効となる問題クラスの特長と、振る舞いの定義法の違いが生み出す性能差についての検証を行っている。実験では、ロボットへの応用でよく見られる問題クラスにおいて、提案手法がNEATのロバスト性を高めることを示している。

第4章「CMA-TWEANN: Sample Efficient Topologically Explorative Neuroevolution」では、CMA-ESを用いた進化型ニューラルネットに対し、トポロジ拡張ルールを付与することで、ロバスト性と高サンプル効率を両立させるCMA-TWEANNと呼ばれるアルゴリズムを提案している。これまでのCMA-ESを用いた進化型ニューラルネットでは、ニューラルネットのトポロジが固定されていることを前提としてサンプル効率を高めており、次元数が途中で変動するようなトポロジ拡張との両立は実現されていなかった。本章では、CMA-ESの特性である探索経験の有効利用を、トポロジ拡張が起きる場合でも可能とするようなアルゴリズムを提案している。実験では、分野の標準的なベンチマークテストにおいて、提案手法がロバスト性と高サンプル効率を両立させることを示している。

第5章「Learning Symbolic Forward Models」では、モデル学習においてシンボリック回帰を用いることで、ロボットの行動計画・制御に適した動力学モデル学習をする手法を提案している。提案手法の有用性は、理想的なロボットの運動方程式が数式として表されること、また数式モデルの計算効率が極めて高いことに依拠している。本章では、学習された動力学モデルは予測精度が高く、さらに計算効率の高さから行動計画において多くの反復が可能となること、そのために、既存の学習法であるガウシアンプロセス回帰やサポートベクター回帰と比べ、高い制御効率を実現することを実証している。

第6章「Conclusion and Future Work」では、本論文の主要な貢献がまとめられている。本論文の主要な貢献の第一は、様々な種類のロボット応用に向けて、それぞれに適した一連の最先端の強化学習アルゴリズムを提案したことである。実世界には多様なロボットの強化学習問題が存在しているため、本論文が提案した一連のアルゴリズムはユーザにとって実問題を解く際の有力な選択肢となる。第二の貢献は、ロボットの強化学習問題における進化計算の有用性を示したことである。特に、アルゴリズムの柔軟性と初期収束に対するロバスト性という進化計算の特長が、強化学習問題において有効に機能することを示している。

以上を要するに、本論文は、ロボットを自動制御するために必要不可欠な技術を考察し、進化計算を活用するアプローチにより、従来手法では実現不可能であった効率性やロバスト性を実現する強化学習手法を提案している。すべての提案手法について、ロボットの強化学習問題により有用性を実証的に示しており、これらは情報理工学分野の今後の発展に寄与・貢献するものである。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。