審 査 の 結 果 の 要 旨

氏　　名　　松尾　凌輔

修士（工学）松尾凌輔提出の論文は「Efficient Reinforcement Learning Using Simulation for Robot Control: Application to Logistics Cart Transportation System （シミュレーションを活用した実ロボット制御のための効率的な強化学習: 台車搬送システムへの適用)」と題し、英文で書かれ、5章から構成されている。

近年、無人航空機や宇宙機、移動ロボットなどにおける高度な自律的制御・意思決定を実現する手段として強化学習に大きな期待が寄せられているが、最適な方策や状態価値関数を学習するためには膨大な試行による訓練データが必要であり、その全てを実環境で収集することはコストや時間、安全性の面で困難であるという問題がある。そのため、実環境の代わりに物理シミュレーションに基づく模擬環境を用いることによって学習効率を改善しリスクを削減しようとする試みが提案されているが、環境間には差異があるため、模擬環境で学習された制御則が実環境で常に有効である保証はない。このような背景を踏まえ、本論文は、模擬環境から入手可能な特権的情報と制御に関する知識を活用することによって、効率的かつ環境間差異に対して頑強な強化学習法を提案している。

第１章は序論であり、近年、未知環境におけるロボットや無人航空機などの自律的意思決定・制御に深層強化学習が有望視されている背景と現状を述べている。強化学習は未知環境における試行錯誤的な行動経験から最適な行動則を学習する枠組みであるが、実環境において膨大な経験すなわち訓練データを収集することはコスト・リスクが非常に高く非現実的である。そのため、物理シミュレーションによる模擬環境を用いて訓練を行うことが提案されているが、実環境を完全に再現するものではないため、模擬環境で学習された制御則が実環境でも有効であるとは限らないという問題が依然残っている。本章では、また、本研究の基盤となっている連続的制御のための強化学習、および、ロボティクスにおける物理シミュレーションについて概説している。

第２章では、台車搬送システムを主な題材として、物理シミュレーション上

での分散的残差強化学習による効率的な制御則の学習法を提案している。既存の制御則を事前知識として活用することで、効率的な学習と大きな性能改善が達成されることを実験で示すとともに、模擬環境で獲得された制御方策が実環境における台車搬送に転移可能であることを実証している。

第３章では、実環境でのロボット制御において大きな問題である遅れフィードバックを解決する強化学習法を提案している。この方法では、訓練時には模擬環境から遅延のない観測と報酬を得られることに着目し、それらの特権的情報を活用することによって効率的な学習を行いつつ、テスト時および実環境への転移時には遅延があっても適用可能な制御則が獲得される。台車搬送以外にも様々なタスクで検証を行い、その有効性と従来手法に対する優位性を明らかにしている。

第４章では、実環境と模擬環境との不整合によって起こる問題を解決するためのドメインランダマイゼーション法を提案している。これはランダム化された模擬環境で訓練を行うことによって実環境との差異に対して頑強な制御則を学習する方法である。提案手法では、特権的情報および観測・行動履歴から抽出された時系列特徴を利用し、タスクにとって本質的な状態表現を学習することによって、実環境を理想的なマルコフ決定過程(MDP)に接近させる方法を明らかにし、その有効性を実証している。

第５章は結論であり、本研究の成果をまとめ、今後の課題を議論している。

以上要するに、本論文は、物理シミュレーションによる効率的な訓練方法と学習された制御則の実環境への頑強な転移方法を明らかにするものであり、その成果は航空宇宙工学上およびロボット工学上貢献するところが大きい。

よって本論文は博士（工学）の学位請求論文として合格と認められる。