

論文内容の要旨

博士論文題目: Domain-Adaptive Robust Training and Deployment in Real-World Noisy Data-Driven Robotics Applications

(実世界のノイジーデータ駆動ロボット応用におけるドメイン適応型ロバスト学習と展開)

氏名: Iboudo Wendyam Eric Lionel

ロボット工学と機械学習の統合は、ロボットを従来のプログラムされた機械から、実世界で活動するデータ駆動型の学習者へと進化させる、変革的な転換を意味する。この移行は、主に不完全な実世界のデータから学習するロボットの能力と、困難で動的な環境に適応する能力を中心に、新たな一連の課題をもたらす。機械学習による実世界のロボット応用では、機械学習モデルの訓練、評価、適用という3つの重要な段階を含む。訓練フェーズの有効性はデータセットの質に大きく左右され、ノイズの多いデータや最適でないデータも頻繁に含まれる。同様に、適用フェーズにおけるロボットの性能は、訓練や評価で経験した環境とは異なるターゲット環境への適応性にかかっている。

本研究では、学習フェーズの課題に対処するために、適応的なロバスト性を持つ2つの新しい確率的勾配降下アルゴリズム (At-momentum と AdaTerm) を提案し、学習プロセス中に破損した勾配の影響を検出して低減することで、ロボットがヘビーテールのデータセットに対しても適切な学習を可能にする。同様に、適用フェーズでは、擬似多目的マルコフ決定過程 (PMOMDP) と呼ばれる新しい強化学習のフレームワークを提案し、sDRSAC、cMDSAC、eMDSAC、uMDSAC という様々なアルゴリズムを提示する。データセット中に未知の外れ値が含まれる実用的な設定において、適応的ロバスト性の有効性を調査するために、模倣学習エージェントに提供されるデモに最適行動と非最適行動の両方が含まれるロバスト行動クローニング (RBC) タスクで At-momentum 最適化アルゴリズムを評価した。この評価では、2台のロボット、QbChain と D'Claw を用いた。提案アルゴリズムでは適応的ロバスト性により、学習データセット中の非最適行動の割合が変化しても、一定の性能を維持できることが示された。一方、非ロバスト最適化では性能が低下し、固定ロバスト最適化では低い性能から始まり、非最適行動の数に応じて性能が上昇することも確認した。また、At-momentum と AdaTerm の両方について

て、画像分類やロボットのダイナミクス予測から強化学習やポリシーの蒸留まで、4つの主要な機械学習タスクで評価したところ、AdaTerm は全てのタスクで一貫して良好な性能を示し、At-momentum が僅差でこれに続く結果となった。

さらに、PMOMDP フレームワークについて、まず明示的なスカラライゼーションが領域ランダム化に与える影響（sDRSAC と DRSAC の比較）を検討し、sDRSAC が特定の環境において、より良い政策と、より良いシミュレーションから実データへの転移性能につながることを明らかにした。次に、三つの MDSAC から生成された不確実性を考慮したポリシーを Mujoco 環境で評価し、様々なシミュレーションパラメータと不確実性レベルの下での性能を測定した。最後に、シミュレーションと実世界の D'Claw ロボット評価の両方において、不確実性を減少させることによる性能向上を評価するために、不確実性を考慮したポリシーを確率的システム同定戦略と組み合わせた手法を評価し、実験結果から有効性を確認した。

氏名	Ilboudo Wendyam Eric Lionel
----	--------------------------------

(論文審査結果の要旨)

本論文は、機械学習の実世界ロボット応用について検討している。機械学習は、訓練、評価、適用、という3つのフェーズを通過するが、訓練フェーズの有効性は訓練データセットの品質に大きく左右される。実世界ロボットタスクにおいては、センサノイズやデータ欠損などの様々な要因により、データセットの品質を担保することは困難である。同様に、適用フェーズの有効性は、訓練や評価で経験した環境とは大きく異なる可能性のあるターゲット環境への適応性に依存し、環境が多様に変化し得る実世界ロボットタスクでは重要な課題となる。本論文では、訓練フェーズと適用フェーズに対する上記の課題に対処する手法を研究し、実用的なアプローチを示している。

学習フェーズの課題に対処するために、適応的なロバスト性を持つ2つの新しい確率的勾配降下アルゴリズム (At-momentum と AdaTerm) を提案した。ノイジーデータの影響を勾配推定のレベルで対処するアイデアの新規性および実用性は高く評価できるものとする。次に、適用フェーズの課題に対処するために、訓練や評価で経験した環境とは大きく異なる可能性のあるターゲット環境への適応性を実現するドメインランダム化強化学習フレームワークとして、擬似多目的マルコフ決定過程 (PMOMDP) と複数の実装アルゴリズムを提案した。ドメインランダム化強化学習を、擬似的な多目的マルコフ決定過程として捉えるアプローチの独自性は高い。また、数理的に整備された MOMDP の知見を活用することで、ドメインランダム化強化学習の新たな可能性が開拓されており、今後の発展が期待できる。

公聴会では、ノイジーデータとデータ拡張の違い、ロバスト性の定義、ユニバーサル方策や多目的性の関係、提案法の限界や拡張性など、多岐にわたる項目について審査委員との質疑が行われた。その後の最終審査において、すべての質問への回答と論文修正について説明が行われ、適切な対応がとられたことを審査員全員で確認した。

本論文は、機械学習の実世界ロボット応用に向けて、訓練時のロバスト性と、適用時の環境適応性の2つの観点に着目した独自のアプローチと複数の実現アルゴリズムを提案した。シミュレーションおよび実ロボットを用いたタスクなどに対する評価実験を実施していることに鑑み、新規性および有用性の観点から一定の学術的意義があるものと評価できる。よって、本論文は博士(工学)の学位論文として価値あるものと認める。