FY2025/2025年度

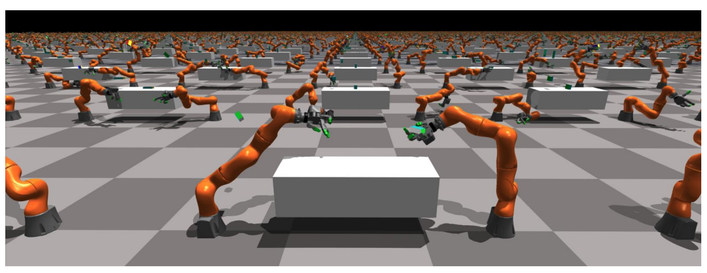
Robot Learning Team
Takayuki Osa
ロボットラーニングチーム　長　隆之

大規模並列環境を活用した強化学習アルゴリズム

N. Shitanda, M. Omura, T. Harada, T. Osa. Rethinking Policy Diversity in Ensemble Policy Gradient in Large-Scale Reinforcement Learning. ICLR 2026 (accepted, to appear).

背景

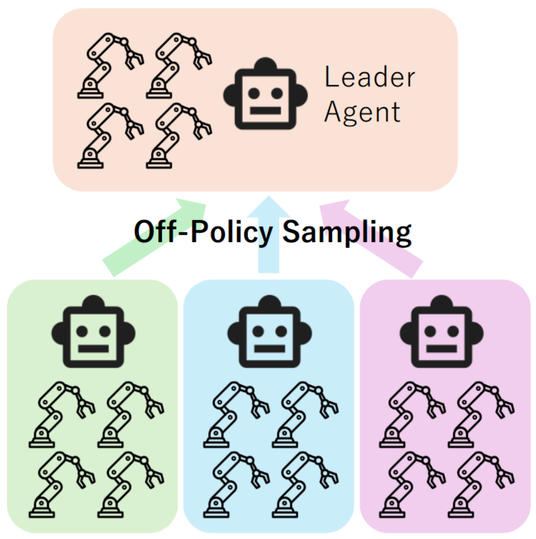
- GPUを用いた数万規模の並列環境の利用が普及
- 数万規模の並列環境を用いることを前提としたアルゴリズムの研究は限られている



既存のアルゴリズム

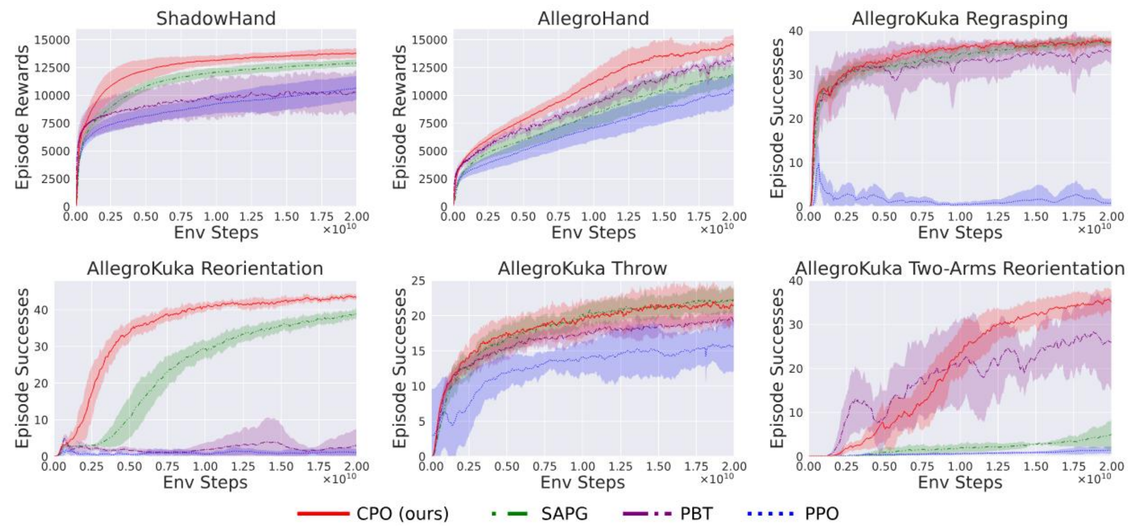
メインの方策(leader)の周辺に若干異なる方策(follower)を配置することで、効果的にデータを収集する[Singla et al., ICML 2024.]

leader方策とfollower方策の位置関係に関する議論は十分でなかった



本研究の貢献

- PPOをベースとした方策更新において、leader方策とfollower方策の密度比が推定される方策勾配のバイアスにどのような影響を与えるかを数理的に明らかにした。
- leader方策とfollower方策のKL情報量を明示的に制御することで学習効率を高めるアルゴリズムを開発し、既存手法より高い性能を示すことを実験的に示した。



クロス・エンボディメント型オフライン強化学習アルゴリズム

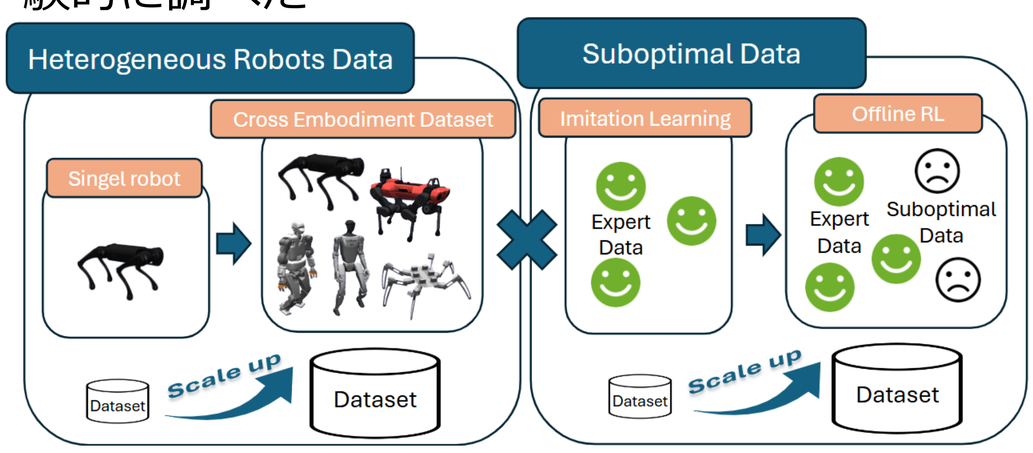
H. Abe, T. Osa, Y. Mukuta, T. Harada. Cross-Embodiment Offline Reinforcement Learning for Heterogeneous Robot Datasets. ICLR 2026 (accepted, to appear).

背景

- 事前に収集されたデータセットのみで学習を行うオフライン強化学習はデータの利用効率を高める可能性を持つ
- ロボットにおける動作の学習では、ロボットの身体性（embodiment）が大きな影響を持つ
- 教師あり学習の文脈においては、多様なロボットからのデータの活用に関する研究はあるが、オフライン強化学習については、ほとんど研究がない

本研究の貢献

- 異なる身体形状を持つロボットの動作データを一つのデータセットして構築し、クロス・エンボディメント型オフライン強化学習について実験的に調べた

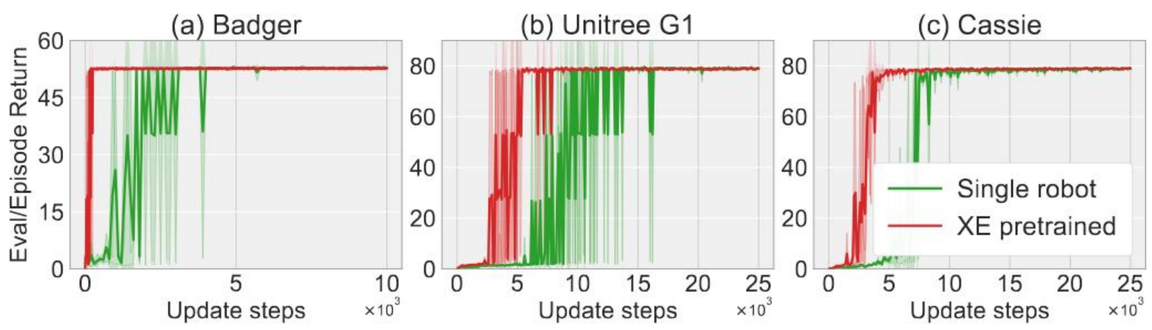


本研究での設定

- Unified Robot Morphology Architecture (URMA)を用いた、異なる次元の観測および行動空間に対応可能なモデル[Bohlinger et al., CoRL 2024]を使用した
- Implicit Q-learning [Kostrikov et al., ICLR 2022]をベースアルゴリズムとして検証した

クロス・エンボディメント型オフライン強化学習の効果

- 異なるエンボディメントのデータを用いたオフライン強化学習によって、その後の異なるロボットでの動作の学習を早めることができた



- 異なるロボット間で、正の転移と負の転移が起きていることが確認された
- エンボディメント間の距離を定量化し、グルーピングすることで、学習効率が高まることを実験的に示した