摘要

具身智能是通向通用人工智能的关键路径，要求智能体能够适应动态环境，应对多样长程任务，兼容异构本体，并在未见的任务和场景中展现泛化能力。强化学习作为实现具身智能的核心方法论，面临样本效率低与泛化能力差的双重瓶颈，根源在于具身环境马尔可夫决策过程在状态、时间和动作三个维度上的复杂性引发的维度灾难。针对这一挑战，本文提出数据驱动的决策空间抽象方法论，利用离线数据和人类演示中的先验知识，将高维决策空间分别映射至紧凑的状态表征、分层的时域抽象和低维的动作抽象表示，通过有效的抽象提升样本效率与泛化能力。主要贡献如下：

（1）**基于状态表征抽象的动态环境泛化。**针对离线元强化学习中行为策略特征与任务信息耦合引发的分布外上下文泛化问题，提出一种基于对比学习的任务表征方法。该方法通过分层任务编码器对单步转移样本进行条件互信息最大化的表征学习，有效解耦行为策略特征，提取鲁棒的任务表征。实验表明该方法在多个基准环境的分布外测试中展现出优于已有方法的鲁棒性。

（2）**基于时域抽象的长程任务泛化。**提出两项递进的工作以解决长程具身任务的高效学习和快速适应泛化问题。第一项工作提出在大规模数据集上预训练目标条件策略和目标先验模型，并利用目标空间聚类实现分层强化学习，在机器人操作环境和开放世界游戏 Minecraft 中达到最优样本效率。第二项工作在此基础上提出预训练多目标 Transformer 策略和提示词优化方法，进一步降低了智能体在未见环境中的探索成本。

（3）**基于动作抽象的复杂物理交互泛化。**面向具身智能体在复杂物理交互中的跨物体、场景与跨本体等泛化挑战，以机器人灵巧手通用抓取为主要场景展开研究。第一项工作将长程高维控制问题抽象为对演示轨迹的单步编辑，在数千个物体的仿真评测上达到最优成功率，并在真实机器人上对 110 个未见物体和复杂摆放的场景实现高成功率的零样本泛化。第二项工作从人类抓取数据中提取手势主成分构建低维动作流形，统一异构灵巧手的动作空间，支撑跨本体的协同训练与零样本泛化。

本研究从状态、时间、动作三个维度系统性地验证了数据驱动的决策空间抽象是降低具身强化学习探索代价、提升泛化能力的有效途径，在连续控制、开放世界游戏和机器人操作等多类环境中取得了优于已有方法的性能，为构建通用具身智能体提供了方法论指导与算法支撑。

关键词：强化学习，具身智能，泛化性，样本效率，决策空间抽象

Decision Space Abstraction for Generalizable Embodied Reinforcement Learning

Haoqi Yuan (Computer Applied Technology)
Directed by Prof. Zongqing Lu

ABSTRACT

Embodied intelligence is a key pathway toward artificial general intelligence, requiring agents to adapt to dynamic environments, handle diverse long-horizon tasks, accommodate heterogeneous embodiments, and generalize to unseen scenarios. As a core methodology for embodied intelligence, reinforcement learning (RL) suffers from low sample efficiency and poor generalization, stemming from the curse of dimensionality in the state, temporal, and action dimensions of the Markov Decision Process in embodied environments. This dissertation proposes a data-driven decision space abstraction methodology that leverages prior knowledge from offline data and human demonstrations to map the high-dimensional decision space into compact state representations, hierarchical temporal abstractions, and low-dimensional action representations, thereby improving sample efficiency and generalization. The main contributions are as follows:

(1) **State representation abstraction for dynamic environment generalization.** To address the entanglement between behavior policy features and task information in offline meta-RL, which leads to generalization failures under out-of-distribution contexts, this work proposes a contrastive learning-based task representation method. A bi-level task encoder learns representations from single-step transitions by maximizing conditional mutual information, effectively disentangling behavior policy features to extract robust task representations. Experiments on multiple benchmarks demonstrate superior robustness over existing methods under out-of-distribution evaluation.

(2) **Temporal abstraction for long-horizon task generalization.** Two progressive works are proposed to enable efficient learning and fast adaptation in long-horizon embodied tasks. The first pre-trains goal-conditioned policies and goal prior models on large-scale task-agnostic datasets and clusters the goal space for hierarchical RL, achieving state-of-the-art sample efficiency in robotic manipulation and the open-world game Minecraft. Building upon this, the

second work pre-trains a multi-goal Transformer policy and introduces prompt optimization, further reducing exploration costs for adapting to unseen environments.

(3) **Action abstraction for generalization in complex physical interactions.** Targeting cross-object, cross-scene, and cross-embodiment generalization challenges, this dissertation studies universal grasping with dexterous hands. The first work abstracts the long-horizon high-dimensional control problem into single-step demonstration editing, achieving state-of-the-art success rates on thousands of simulated objects and high-success-rate zero-shot generalization to 110 unseen objects and complex layouts on a real robot. The second work extracts principal components from human grasp data to construct a low-dimensional action manifold that unifies the action spaces of heterogeneous dexterous hands, enabling cross-embodiment co-training and zero-shot generalization to unseen hands.

This dissertation systematically validates, across the state, temporal, and action dimensions, that data-driven decision space abstraction is an effective approach to reducing exploration costs and enhancing generalization in embodied RL. The proposed methods achieve superior performance in continuous control, open-world games, and robotic manipulation, providing methodological guidance and algorithmic foundations for building general-purpose embodied agents.

KEYWORDS: Reinforcement Learning, Embodied AI, Generalization, Sample Efficiency, Decision Space Abstraction