Cognizing and Imitating Robotic Skills via a Dual Cognition-Action Architecture

通过双重认知-动作架构认知与模仿机器人技能

Extended Abstract

扩展摘要

Zixuan Chen

陈子轩

State Key Laboratory for Novel

新型软件技术国家重点实验室

Software Technology

软件技术

Nanjing University

南京大学

Nanjing, China

中国南京

chenzx@nju.edu.cn
Ze Ji

纪泽

Cardiff University

卡迪夫大学

Cardiff, United Kingdom

英国卡迪夫

jiz1@cardiff.ac.uk
Shuyang Liu

刘书阳

State Key Laboratory for Novel

新型软件技术国家重点实验室

Software Technology

软件技术

Nanjing University

南京大学

Nanjing, China

中国南京

MG20330036@smail.nju.edu.cn
Jing Huo

霍晶

State Key Laboratory for Novel

新型软件技术国家重点实验室

Software Technology

软件技术

Nanjing University

南京大学

Nanjing, China

中国南京

huojing@nju.edu.cn
Yiyu Chen

陈奕宇

State Key Laboratory for Novel

新型软件技术国家重点实验室

Software Technology

软件技术

Nanjing University

南京大学

Nanjing, China

中国南京

yiyuiii@foxmail.com
Yang Gao

高扬

State Key Laboratory for Novel

新型软件技术国家重点实验室

Software Technology

软件技术

Nanjing University

南京大学

Nanjing, China

中国南京

gaoy@nju.edu.cn

Abstract

摘要

Enabling robots to effectively learn and imitate expert skills in long-horizon tasks remains challenging. Hierarchical imitation learning (HIL) approaches have made strides but often fall short in complex scenarios due to their reliance on self-exploration. This paper introduces a novel approach inspired by the human skill acquisition process, proposing a Cognition-Action-based Robotic Skill Imitation Learning (CasIL) framework. CasIL integrates human cognitive priors for task decomposition into a dual-layer architecture, enhancing robots' ability to cognize and imitate essential skills from expert demonstrations. Our experiments across four RLbench tasks demonstrate CasIL's superior performance, robustness, and generalizability in skill imitation compared to related methods.

使机器人能够在长时任务中有效学习和模仿专家技能仍然具有挑战性。层次模仿学习 (Hierarchical Imitation Learning, HIL) 方法虽有所进展, 但由于依赖自我探索, 在复杂场景中常常表现不足。本文提出一种受人类技能习得过程启发的新方法, 提出了基于认知-动作的机器人技能模仿学习 (Cognition-Action-based Robotic Skill Imitation Learning, CasIL) 框架。CasIL 将人类认知先验用于任务分解，构建双层架构，增强机器人从专家示范中认知和模仿关键技能的能力。我们在四个 RLbench 任务上的实验表明，CasIL 在技能模仿的性能、鲁棒性和泛化能力方面均优于相关方法。

KEYWORDS

关键词

Hierarchical imitation learning, Robotic skill imitation, Visual demonstrations

分层模仿学习，机器人技能模仿，视觉示范

ACM Reference Format:

ACM 参考格式:

Zixuan Chen, Ze Ji, Shuyang Liu, Jing Huo, Yiyu Chen, and Yang Gao. 2024. Cognizing and Imitating Robotic Skills via a Dual Cognition-Action Architecture: Extended Abstract. In Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6-10, 2024, IFAAMAS, 3 pages.

陈子轩，纪泽，刘书阳，霍晶，陈奕宇，高扬。2024。通过双重认知-动作架构认知与模仿机器人技能: 扩展摘要。载于第 23 届国际自治代理与多智能体系统会议 (AAMAS 2024) 论文集，新西兰奥克兰，2024 年 5 月 6-10 日，IFAAMAS，3 页。

1 INTRODUCTION

1 引言

To advance robot skill imitation in long-horizon tasks, Hierarchical Imitation Learning (HIL) is recognized for overcoming traditional IL preprocessing challenges [1, 3, 8] . HIL enables robots to learn from expert demonstrations through a two-tier policy structure: acquiring sub-policies for specific task segments at the lower level and overarching strategies for skill transition at the higher level. However, HIL's effectiveness depends on the robustness of its hierarchical structure, with weaknesses leading to subpar imitation. Recognizing the limitations of relying solely on deep learning for hierarchy development in HIL, we draw inspiration from human cognitive processes in skill acquisition. This process emphasizes the dynamic interaction of information processing, task decomposition, decision-making, and refinement, with a significant emphasis on the integration of prior knowledge and observed behaviors through working memory [4, 6, 7] . Building on these principles, we propose the Cognition-Action-based Robotic Skill Imitation Learning (CasIL) framework. CasIL introduces a novel dual cognition-action structure for effective skill imitation in complex tasks, incorporating operators' cognitive priors for enhanced learning efficiency.

为推动机器人在长时序任务中的技能模仿，分层模仿学习 (Hierarchical Imitation Learning, HIL) 因克服传统模仿学习预处理难题而备受认可 $[1,3,8]$ 。HIL 使机器人通过两级策略结构从专家示范中学习: 低层获取特定任务片段的子策略，高层掌握技能转换的总体策略。然而，HIL 的有效性依赖于其分层结构的稳健性，结构薄弱会导致模仿效果不佳。鉴于仅依赖深度学习构建 HIL 层级存在局限，我们借鉴人类技能习得中的认知过程。该过程强调信息处理、任务分解、决策与优化的动态交互，特别注重通过工作记忆整合先验知识与观察行为 $[4,6,7]$ 。基于此，我们提出了基于认知-动作的机器人技能模仿学习 (Cognition-Action-based Robotic Skill Imitation Learning, CasIL) 框架。CasIL 引入新颖的双重认知-动作结构，用于复杂任务中的高效技能模仿，融合操作者的认知先验以提升学习效率。

2 COGNITION-ACTION-BASED SKILL IMITATION LEARNING

2 基于认知-动作的技能模仿学习

In our problem formulation, we model long-horizon task environments as Semi-Markov Decision Processes (SMDP), represented by the tuple $\left(\mathcal{S}, \mathcal{A}, \{I_o, \pi_o, \beta_o\}_{o \in O}, \pi_O\left(o \mid s\right), \mathcal{P}, \mathcal{R}\right)$. Here, $\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}$ are standard MDP components, with the addition of $\{\mathcal{I}_o, \pi_o, \beta_o\}$ for each option o in the option set O . An option, comprising a policy $\pi_o : \mathcal{S} \times \mathcal{A} \rightarrow [0,1]$, a termination condition $\beta_o : \mathcal{S}^+ \rightarrow [0,1]$, and an initiation set $\mathcal{I}_o \subseteq \mathcal{S}$, is valid in state s_t iff $s_t \in \mathcal{I}_o$. The system transitions between options based on the termination condition β_o and the inter-option policy $\pi_O\left(o \mid s\right)$, progressing until the task is completed. The CasIL framework, illustrated in Fig. 1, features three main components. Initially, pre-trained image and text encoders process the visual and textual inputs. Following this, a cognition generator $\mathcal{F} : G \times \mathcal{S} \rightarrow O$ and a policy module $\pi_O : \mathcal{S} \times O \rightarrow \mathcal{A}$ operate in tandem. Here, $O = \left\{o^{\mathbf{1}}, \ldots, o^{\mathbf{K}}\right\}$ denotes a set of K options, with each option representing a sub-task equipped with a specific skill, together forming a skill chain. CasIL works through two phases: 1) Leveraging manually inputted cognitive priors for task decomposition and expert visual demonstrations, the robot constructs its cognition-action framework and skill chain O , guided by the task objectives. 2) The robot then chooses the most appropriate sub-task skill from O , based on its observation history, and learns and implements the relevant policies π_O to accomplish the sub-tasks. Using expert demonstration $\tau^E = \left(G, \{s_t, a_t\}_{t=1}^{T}\right)$ and textual decompositions $\{l_1, \ldots, l_{\mathbf{K}}\}$ based on human cognitive priors, the cognition generator $\mathcal{F}$ aligns states s_t with decompositions $l_{\mathbf{t}}$ to produce essential skills $o^{\mathbf{t}}$ for each division, extending the demonstration into an option-expanded trajectory $\tau^E = \left(G, \{s_t, o_{\mathbf{t}}, a_t\}_{1 \leq t \leq T}^{1 \leq \mathbf{t} \leq \mathbf{K}}\right)$, creating an SMDP structure. The robot selects relevant skills $o^{\mathbf{t}}$ based on the goal and observations, with each skill $o^{\mathbf{t}}$ active for $H\left(\mathbf{t}\right)$ time steps. The policy π then guides actions at each step, depending on the state and the current skill. A CasIL-equipped robot utilizes human cognitive priors to learn and form its cognition of skills from expert demonstrations, focusing on critical decision-making steps. This learning approach enables the robot to adapt its actions based on observed inputs, following a dual learning framework. CasIL's training involves both a high-level cognition generator for skill chain encoding and a low-level action module for skill execution, as illustrated in Fig. 1. The cognition generator aligns task goals with human cognitive priors and expert demonstrations, while the low-level module employs behavior cloning with options. The training objective for a trajectory of length T aims to minimize the loss function:

在我们的问题表述中，我们将长时域任务环境建模为半马尔可夫决策过程 (Semi-Markov Decision Processes，SMDP)，用元组 $\left(\mathcal{S}, \mathcal{A}, \{I_o, \pi_o, \beta_o\}_{o \in O}, \pi_O\left(o \mid s\right), \mathcal{P}, \mathcal{R}\right)$ 表示。这里，$\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}$ 是标准的 MDP 组件，另外为选项集 O 中的每个选项 o 增加了 $\{\mathcal{I}_o, \pi_o, \beta_o\}$。一个选项由策略 $\pi_o : \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$、终止条件 $\beta_o : \mathcal{S}^+ \rightarrow [0, 1]$ 和启动集 $\mathcal{I}_o \subseteq \mathcal{S}$ 组成，当且仅当 $s_t \in \mathcal{I}_o$ 时在状态 s_t 中有效。系统根据终止条件 β_o 和选项间策略 $\pi_O\left(o \mid s\right)$ 在选项之间转换，直到任务完成。CasIL 框架如图 1 所示，包含三个主要部分。首先，预训练的图像和文本编码器处理视觉和文本输入。随后，认知生成器 $\mathcal{F} : G \times \mathcal{S} \rightarrow O$ 和策略模块 $\pi_O : \mathcal{S} \times O \rightarrow \mathcal{A}$ 协同工作。这里，$O = \{o^{\mathbf{1}}, \ldots, o^{\mathbf{K}}\}$ 表示一组 K 选项，每个选项代表一个配备特定技能的子任务，共同构成技能链。CasIL 通过两个阶段工作:1) 利用手动输入的认知先验进行任务分解和专家视觉示范，机器人构建其认知-动作框架和技能链 O，以任务目标为指导。2) 机器人基于观察历史从 O 中选择最合适的子任务技能，学习并执行相关策略 π_O 以完成子任务。通过专家示范 $\tau^E = \left(G, \{s_t, a_t\}_{t=1}^{T}\right)$ 和基于人类认知先验的文本分解 $\{l_1, \ldots, l_{\mathbf{K}}\}$，认知生成器 $\mathcal{F}$ 将状态 s_t 与分解 $l_{\mathbf{t}}$ 对齐，生成每个分段的关键技能 $o^{\mathbf{t}}$，将示范扩展为选项扩展轨迹 $\tau^E = \left(G, \{s_t, o_{\mathbf{t}}, a_t\}_{1 \leq t \leq T}^{1 \leq \mathbf{t} \leq \mathbf{K}}\right)$，构建 SMDP 结构。机器人根据目标和观察选择相关技能 $o^{\mathbf{t}}$，每个技能 $o^{\mathbf{t}}$ 持续 $H\left(\mathbf{t}\right)$ 时间步。策略 π 则根据状态和当前技能指导每一步的动作。配备 CasIL 的机器人利用人类认知先验从专家示范中学习并形成技能认知，聚焦关键决策步骤。该学习方法使机器人能够基于观察输入调整动作，遵循双重学习框架。CasIL 的训练包括用于技能链编码的高层认知生成器和用于技能执行的低层动作模块，如图 1 所示。认知生成器将任务目标与人类认知先验及专家示范对齐，低层模块则采用带选项的行为克隆。长度为 T 的轨迹的训练目标是最小化损失函数:

$$\mathcal{L}_{\text{CasIL}} = \min_{\theta_g, \theta_p} \sum_{t=1}^{T} \left(-\varepsilon \log \mathcal{F}_\xi \left(o^{\mathbf{t}} \mid G, \{s_\kappa\}_{\kappa=1}^{\kappa=\sum H(\mathbf{t})}, \{o^\kappa\}_{\kappa=1}^{\mathbf{t}-1} \right) \right. \tag{1}$$

$$\left. \log \pi_\theta \left(\widehat{a}_t \middle| \widehat{G}, \{\widehat{s}_\kappa\}_{\kappa=1}^{\kappa=\sum H(\mathbf{t})}, o^{\mathbf{t}}) \middle| \right) \right. .$$

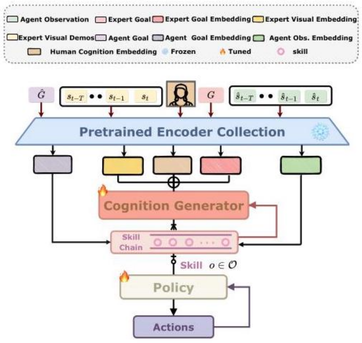

Figure 1: The workflow of CasIL.

图 1:CasIL 的工作流程。

where ξ and θ represent the weights for the cognition generator and policy module, respectively, and ε adjusts the cognitive generation loss. The symbols o, s, a , and G denote skills, observations, actions, and task goals, consistent with the initial definitions.

其中 ξ 和 θ 分别表示认知生成器和策略模块的权重，ε 调整认知生成损失。符号 o, s, a 和 G 分别表示技能、观察、动作和任务目标，与初始定义一致。

3 EXPERIMENTS

3 实验

Our experiments on RLBench [2] assess the methods in robotic arm manipulation tasks across four increasingly complex settings, each with 100 demonstration trajectories for training. Each setup involves a 6-DOF robotic arm with a gripper: ToiletSeatDown: The task is to lower the toilet lid onto the seat within 200 time steps. PutRubbishInBin: The robot must pick up and dispose of rubbish into a bin within 250 time steps. PlayJenga: The robot aims to remove a protruding block from a Jenga tower without toppling it, within 300 time steps. InsertUsbInComputer: The robot needs to pick up a USB stick and insert it into a USB port within 400 time steps. We assess models using success rates' mean and standard deviation in 80 randomized scenarios. Comparative methods include: 1) Supervised Behavioral Cloning (BC) [5]: Lacks hierarchical structure and cognitive inputs. 2) Hierarchical Behavioral Cloning (H-BC) [9]: Uses an option-based architecture without human cognitive priors.

我们在 RLBench[2] 上的实验评估了机器人手臂操作任务中的方法，涵盖四个复杂度逐渐增加的设置，每个设置包含 100 条示范轨迹用于训练。每个环境均配备一个带夹持器的 6 自由度机器人手臂:ToiletSeatDown: 任务是在 200 个时间步内将马桶盖放下至座圈上。

Robotic Arm Manipulation				
	ToiletSeatDown	PutRubbishInBin	PlayJenga	InsertUsbInComputer
BC	93.7 ± 4.3	74.4 ± 3.7	21.5 ± 8.8	00.0 ± 0.0
H-BC	98.5 ± 1.5	85.2 ± 5.9	33.6 ± 7.9	10.6 ± 1.8
Option-GAIL	99.0 ± 1.0	81.4 ± 9.6	48.2 ± 9.2	23.3 ± 5.5
CasIL w/o Cognition	99.4 ± 0.6	89.6 ± 9.4	53.1 ± 8.3	26.4 ± 4.1
CasIL (ours)	100.0 ± 0.0	98.4 ± 1.6	82.4 ± 3.5	57.6 ± 2.4

机器人手臂操作				
	放下马桶盖	将垃圾放入垃圾桶	玩积木叠叠乐	将 USB 插入电脑
BC	93.7 ± 4.3	74.4 ± 3.7	21.5 ± 8.8	00.0 ± 0.0
H-BC	98.5 ± 1.5	85.2 ± 5.9	33.6 ± 7.9	10.6 ± 1.8
Option-GAIL	99.0 ± 1.0	81.4 ± 9.6	48.2 ± 9.2	23.3 ± 5.5
无认知的 CasIL	99.4 ± 0.6	89.6 ± 9.4	53.1 ± 8.3	26.4 ± 4.1
CasIL(我们的)	100.0 ± 0.0	98.4 ± 1.6	82.4 ± 3.5	57.6 ± 2.4

Table 1: Comparison of test results under four RLBench tasks.

表 1: 四个 RLBench 任务下测试结果的比较。

3) Option-GAIL [3]: Hierarchical, includes self-exploration but omits human cognitive guidance. 4) CasIL w/o Cognition: CasIL variant without the cognition generator to highlight the importance of cognitive modeling. Test results in Table 1 reveal that all methods, including our CasIL, perform well in the simple ToiletSeatDown task, with CasIL achieving a 100% success rate across all test tasks. However, as task complexity increases (with more objects, longer periods and reduced stability), BC's success rate in skill imitation plummets, dropping to 0% in all Inser-tUsbInComputer test tasks. Baselines like H-BC and Option-GAIL, which lack the guidance of human cognitive priors, significantly lag behind CasIL in skill imitation. Similarly, CasIL w/o Cognition struggles with stable manipulation due to the absence of ongoing text-image alignment training. The performance of Option-GAIL, in particular, indicates that a one-step option architecture based solely on agent self-exploration fails to ensure stable skill imitation in long-horizon tasks.

步骤。PutRubbishInBin: 机器人必须在 250 个时间步内拾取并将垃圾丢入垃圾桶。PlayJenga: 机器人目标是在 300 个时间步内从积木塔中移除突出的一块积木且不使其倒塌。InsertUsbInComputer: 机器人需要在 400 个时间步内拾取 USB 并将其插入 USB 接口。我们通过 80 个随机场景中成功率的均值和标准差评估模型。比较方法包括:1) 监督行为克隆 (BC)[5]: 缺乏层次结构和认知输入。2) 层次行为克隆 (H-BC)[9]: 采用基于选项的架构但无人工认知先验。3)Option-GAIL [3]: 层次化，包含自我探索但缺乏人工认知指导。4) 无认知的 CasIL:CasIL 变体，去除认知生成器以突出认知建模的重要性。表 1 中的测试结果显示，所有方法包括我们的 CasIL 在简单的 ToiletSeatDown 任务中表现良好，CasIL 在所有测试任务中均达到 100% 的成功率。然而，随着任务复杂度增加 (更多对象、更长时间和稳定性降低)，BC 在技能模仿中的成功率急剧下降，在所有 InsertUsbInComputer 测试任务中降至 0%。缺乏人工认知先验指导的基线方法如 H-BC 和 Option-GAIL 在技能模仿上明显落后于 CasIL。同样，缺乏持续文本-图像对齐训练的无认知 CasIL 在稳定操作上表现不佳。Option-GAIL 的表现尤其表明，单步选项架构仅依赖代理自我探索，无法保证长时任务中技能模仿的稳定性。

4 CONCLUSION

4 结论

We present CasIL, a framework for robot skill imitation using a dual cognition-action architecture. The framework utilizes a text-image-aligned skill chain that is derived from visual expert demonstrations and references human cognitive priors with manual input. This design facilitates robots in cognizing and imitating critical skills for long-horizon tasks. Experimental results show that CasIL improves robot skill imitation performance in long-horizon tasks. Future directions include further enriching cognitive priors and extending the task applicability of CasIL.

我们提出了 CasIL，一种基于双重认知-动作架构的机器人技能模仿框架。该框架利用从视觉专家示范中提取并与文本图像对齐的技能链，结合人工输入的人类认知先验。此设计促进机器人认知并模仿长时任务中的关键技能。实验结果表明，CasIL 提升了机器人在长时任务中的技能模仿表现。未来工作包括进一步丰富认知先验并扩展 CasIL 的任务适用性。

ACKNOWLEDGMENTS

致谢

This work was supported in part by the Science and Technology Innovation 2030 New Generation Artificial Intelligence Major Project under Grant 2021ZD0113303; in part by the National Natural Science Foundation of China under Grant 62192783, Grant 62276128; in part by the Collaborative Innovation Center of Novel Software Technology and Industrialization. REFERENCES

本工作部分由“科技创新 2030 新一代人工智能重大项目”(项目编号 2021ZD0113303) 资助；部分由国家自然科学基金 (项目编号 62192783，62276128) 资助；部分由新型软件技术与产业化协同创新中心支持。参考文献

[1] Jiayu Chen, Tian Lan, and Vaneet Aggarwal. 2023. Option-Aware Adversarial Inverse Reinforcement Learning for Robotic Control. In IEEE International Conference on Robotics and Automation, ICRA 2023, London, UK, May 29 - June 2, 2023. IEEE, 5902-5908.

陈佳宇，兰天，Vaneet Aggarwal. 2023. 面向机器人控制的选项感知对抗逆强化学习. IEEE 国际机器人与自动化会议，ICRA 2023，英国伦敦，2023 年 5 月 29 日-6 月 2 日. IEEE, 5902-5908.

[2] Stephen James, Zicong Ma, David Rovick Arrojo, and Andrew J Davison. 2020. Rlbench: The robot learning benchmark & learning environment. IEEE Robotics and Automation Letters 5, 2 (2020), 3019-3026.

Stephen James, Zicong Ma, David Rovick Arrojo, Andrew J Davison. 2020. RLBench: 机器人学习基准与学习环境. IEEE 机器人与自动化快报 5 卷 2 期 (2020), 3019-3026.

[3] Mingxuan Jing, Wenbing Huang, Fuchun Sun, Xiaojian Ma, Tao Kong, Chuang Gan, and Lei Li. 2021. Adversarial option-aware hierarchical imitation learning. In International Conference on Machine Learning. PMLR, 5097-5106.

景明轩，黄文兵，孙福春，马晓健，孔涛，甘闯，李磊. 2021. 对抗选项感知层次模仿学习. 国际机器学习会议. PMLR, 5097-5106.

[4] Andrew N Meltzoff and Rebecca A Williamson. 2013. Imitation: Social, cognitive, and theoretical perspectives. (2013).

Andrew N Meltzoff, Rebecca A Williamson. 2013. 模仿: 社会、认知与理论视角. (2013).

[5] Dean A Pomerleau. 1988. Alvinn: An autonomous land vehicle in a neural network. Advances in neural information processing systems 1 (1988).

Dean A Pomerleau. 1988. ALVINN: 基于神经网络的自主陆地车辆. 神经信息处理系统进展 1 (1988).

[6] Yingxu Wang. 2007. On the Cognitive Processes of Human Perception with Emotions, Motivations, and Attitudes. Int. J. Cogn. Informatics Nat. Intell. 1, 4 (2007), 1-13.

王英旭. 2007. 关于带有情感、动机和态度的人类感知认知过程. 国际认知信息学与自然智能杂志 1 卷 4 期 (2007), 1-13.

[7] Yingxu Wang and Vincent Chiew. 2010. On the cognitive process of human problem solving. Cogn. Syst. Res. 11, 1 (2010), 81-92.

王英旭，Vincent Chiew. 2010. 关于人类问题解决的认知过程. 认知系统研究 11 卷 1 期 (2010), 81-92.

[8] Dandan Zhang, Qiang Li, Yu Zheng, Lei Wei, Dongsheng Zhang, and Zhengyou Zhang. 2021. Explainable hierarchical imitation learning for robotic drink pouring. IEEE Transactions on Automation Science and Engineering 19, 4 (2021), 3871-3887.

张丹丹, 李强, 郑宇, 魏磊, 张东升, 张正友. 2021. 可解释的层次模仿学习用于机器人倒饮料. IEEE 自动化科学与工程汇刊 19 卷 4 期 (2021), 3871-3887.

[9] Zhiyu Zhang and Ioannis Paschalidis. 2021. Provable hierarchical imitation learning via em. In International Conference on Artificial Intelligence and Statistics. PMLR, 883-891.

张志宇 (Zhiyu Zhang) 和伊奥安尼斯·帕斯卡利迪斯 (Ioannis Paschalidis)。2021 年。通过期望最大化 (EM) 实现可证明的分层模仿学习。载于国际人工智能与统计会议 (International Conference on Artificial Intelligence and Statistics)。PMLR，883-891 页。