Active Fine-tuning of Multi-task Policies

Marco Bagatella [1 2], Jonas Hübotter [1], Georg Martius [2 3], Andreas Krause [1]

[1] ETH Zürich [2] MPI IS Tübingen [3] University of Tübingen

ETH zürich · Learning & Adaptive Systems · Eberhard Karls Universität Tübingen · Max Planck Institute for Intelligent Systems

Background

- Pre-trained generalist policies are gaining relevance in robot learning.
- When fine-tuning on a single task, we should simply collect as many demonstrations as we can afford.
- We study **multi-task** fine-tuning, which presents the problem of dynamically allocating demonstrations to each task.

Algorithm

Criterion:

maximize the expected information gain about the expert policy over its target occupancy

$$c_n = \underbrace{\operatorname*{argmax}_{c' \in \mathcal{C}}}_{\text{task set}} \mathbb{E} \sum_{t=0}^{H-1} \mathcal{I}(\boldsymbol{\pi}(s_t, c), \boldsymbol{\tau}(c') | c_{1:n-1}, \underbrace{\tau_{1:n-1}}_{\text{previous demos}})$$

$$\text{with } c \sim \underbrace{\mu_c}_{\text{target task distribution}}, (s_0, \dots) \sim \boldsymbol{\tau}(c), \tau_{1:n-1} \sim \boldsymbol{\tau}(c_{1:n-1})$$

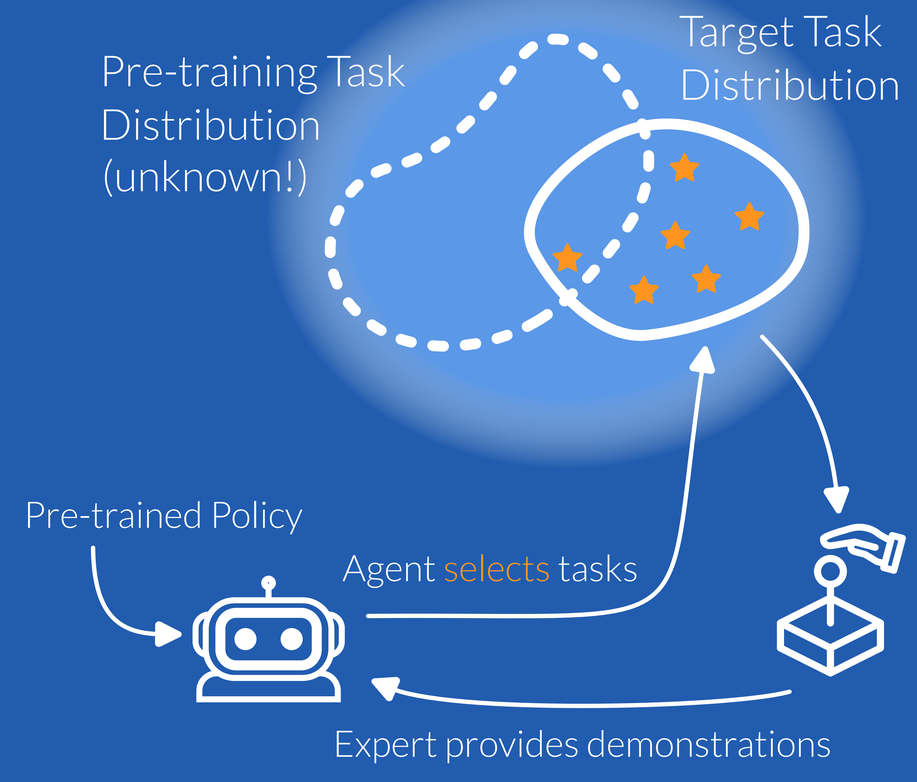
Guarantee:

Under regularity assumptions, the policy achieves expert performance.

Practical ingredients:

1. GP approximation of NNs to estimate mutual information
2. importance sampling to estimate occupancies for unseen tasks
3. prior network to mitigate catastrophic forgetting

We propose a principled criterion to actively decide which tasks should be demonstrated and how often when fine-tuning a pre-trained policy.



Experiments