Robotics, Artificial Intelligence and Embedded Systems

Human Brain Project

MA Proposal: Context-based Meta-Reinforcement Learning with Bayesian Nonparametric Models

Background:

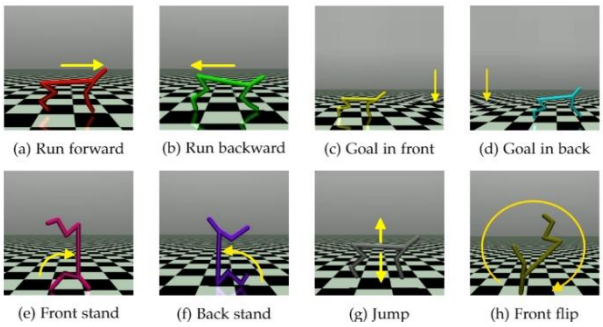
Are you curious about how real research in a newly emerging field of machine learning works? Do you want to contribute to a potential publication on an AI conference? Then look no further.

As humans, we naturally generalize our prior experience to acquire new skills. For example, you may find it easier to learn snowboarding if you already know how to ski. Although reinforcement learning (RL) can surpass humans in learning individual tasks such as Go and Starcraft, there is no mechanism for traditional RL agents to transfer knowledge from similar tasks to facilitate the learning of new tasks. It can be problematic in robotic settings, where it is time-wasting to collect hundreds of thousands of interactions on each individual task with a robot, and the data collection process can cause substantial wear on the hardware. Meta-reinforcement learning (Meta-RL) has been proposed to overcome this shortcoming. Recent meta-RL algorithms such as PEARL [1] train on a distribution of partially observed Markov decision processes (POMDPs) and update a belief state over the task using the evidence collected over multiple trajectories. At test time, these meta-RL agents can perform previously unseen tasks with drastically improved efficiency.

Goal and Methods:

Combining The goal of this project is to design and train meta-RL agents that can approach human-level adaptability and continually learn many tasks over a lifetime. For this, you can leverage the state-of-the-art meta-RL algorithms developed under our chair and extend them with your ideas. You will evaluate your meta-RL agents on the simulated benchmark of Meta-World [2], where the agents learn to perform 50 distinct robotic manipulation tasks such as opening the drawer or pushing a mug under a coffee machine. You may work in the following directions:

1. Our existing algorithm incorporates Bayesian non-parametric (BNP) models to discover the latent task structure during meta-training. BNP models have been used to reconcile continual learning and on-policy meta-RL (MAML) [3]. You can take on the challenge of combining the more efficient off-policy meta-RL and continual learning.
2. To make meta-RL more efficient, we want to effectively share data between different tasks. One possibility is to relabel the experiences collected for one task to another task under certain conditions. This idea of hindsight experience replay has been extended to meta-learning with narrow task distribution [4]. You may work on how to utilize experience replay for a broad task distribution such as Meta-World.



Data Type: Self-generated data in robotic scenarios.

Supervisor: Prof. Alois Knoll;
Advisors:Dr. Zhenshan Bing: bing@in.tum.de,
Lehrstuhl für Robotik, Künstliche Intelligenz und Echtzeitsysteme,
Fakultät für Informatik, Technische Universität München

Related Readings:

[1] Rakelly, Kate, et al. "Efficient off-policy meta-reinforcement learning via probabilistic context variables." *International conference on machine learning*. PMLR, 2019.APA
[2] Meta-World. https://meta-world.github.io/
[3] Jerfel, Ghassen, et al. "Reconciling meta-learning and continual learning with online mixtures of tasks." *Advances in Neural Information Processing Systems* 32 (2019).
[4] Wan, Michael, Jian Peng, and Tanmay Gangwani. "Hindsight Foresight Relabeling for Meta-Reinforcement Learning." *arXiv preprint arXiv:2109.09031* (2021).

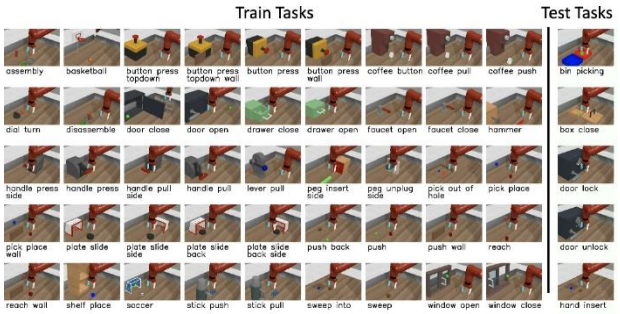


Fig. 1: Cheeta-eight and MetaWorld benchmark for Meta-RL