Robotics, Artificial Intelligence and Embedded Systems

Human Brain Project

MA Proposal: Exploiting DATA Symmetries in Context-Based Meta-Reinforcement Learning

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.

In contrast to supervised learning, which uses examples of how to solve a task correctly, reinforcement learning (RL) only requires a reward that the agent tries to maximize. This powerful approach enabled major breakthroughs like [solving Rubik's cube with a robot hand] and beating humans in games like [Dota 2] or [Go] over the last years. However, these agents are highly specialized and need huge amounts of training experience to solve a task, which may not be feasible–especially in robotic environments. In contrast, the meta-RL setting exposes algorithms to a wide variety of different tasks during training such that they should quickly adapt to new tasks during testing. For this purpose, in inference-based meta-RL a separate encoder is learned in an unsupervised manner which generates a task representation that is then passed to the policy along with the other observations. If, for instance, the different tasks would just be reaching different goal positions, the task representation might be the goal position. This technique has many advantages such as high sample efficiency. To allow optimal behaviour even if the agent cannot know the exact goal yet, a probability distribution over possible specifications can be given to the policy instead. A different concept in AI research which has proven highly effective in computer vision is equivariant neural networks. The idea is simple: in many cases, some transformations of the input should always result in some (potentially different) transformation of the output. For instance, if a segmentation map is generated from an image, categorizing each pixel in cat or non-cat, then rotating or mirroring the input should just rotate or mirror the segmentation map accordingly. By designing the network in a way that ensures this property, it becomes more sample efficient as each training sample also provides information about the desired result in the rotated or mirrored case.

Goal and Methods:

Combining both ideas yields an **equivariant inference-based meta-RL** algorithm with numerous desirable properties. On top of improved sample efficiency, it provides a guarantee for generalization to tasks given by the equivariance transformation. For instance, even if an agent is only trained on tasks on its right in a reflection-equivariant setting, it will reach goals on the left equally well during testing. Given a detailed description of the theoretical background along with a working code base that has been successfully tested on a simple toy environment, the goal of this project is to apply the described concept to new domains. This involves identifying equivariances in different settings, implementing them if they are not present in the code base yet, running experiments and tuning hyperparameters. It may also include optimizing the existing implementation to improve feasibility of complex examples. One important environment that should be evaluated is the MetaWorld benchmark in which a robot arm is tasked to fulfil various different objectives like reaching positions, picking up objects and opening doors. Performing these tasks on the right-hand side is symmetric to performing them on the left-hand side.

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

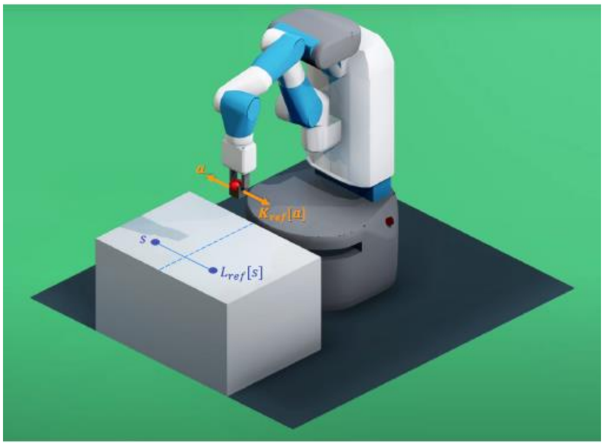

Fig. 1: A reflection equivariant environment with a robot arm

Related Readings:

[1] Efficient off-policy meta-reinforcement learning via probabilistic context variables. *ICML 2019*.

[2] MDP Homomorphic Networks: Group Symmetries in Reinforcement Learning.*Neurips* 2020