Master/Bachelor Thesis – Semester Project

Reinforcement Learning via Hindsight Goal Generation (HGG)

Background

Reinforcement Learning (**RL**) help an agent learn optimal behavior through trial-and-error interactions with an environment. One of the challenges in RL is to design appropriate reward function which involves multiple factors and engineering labors. Therefore, binary reward is used for developing RL algorithms. Recently, Hindsight Experience Replay (**HER**) has been proposed to allow off-policy RL algorithms to perform effective learning in solving goal-based tasks with sparse/binary rewards, such as the manipulation of robotic arms [1]. HER takes advantage of failed trajectories by replacing desired goals with the achieved goals. However, it cannot solve the tasks If desired goal is far away from the initial states. To address this problem, Hindsight Goal Generation (**HGG**) generates intermediate goals which are easy to achieve in short term and lead agent to the desired goals in the long term [2]. Even if the tasks can be accomplished with these goals sampling methods, we need to come up with different efficient exploration methods (Replay Prioritization Methods) to speed up the learning rate. Some proposed replay prioritization methods are Energy Based Prioritization [3] which prioritize samples according to Potential + Kinetic energy of the object, Maximum Entropy-Based Prioritization [4] which prioritize samples based on the trajectory entropy. However, there are still different prioritization methods possible using different kind of sensors such as force sensors [5].

Your Tasks

In this thesis, your task will be learning state-of-the-art knowledge of reinforcement learning and HER, HGG and then develop more advanced algorithms compared vanilla HGG. To be specific:

1. You will first learn basic knowledge of reinforcement learning.
2. You will reproduce the results from HER and other related research results. By doing this, you will have a deep understanding of HER and the state-of-the-art research results.
3. You will explore different replay prioritization methods using Mujoco simulation.

Requirement

- High self-motivation;
- Six month working time;
- Experiences or knowledge from related courses
- Python programming experiences.

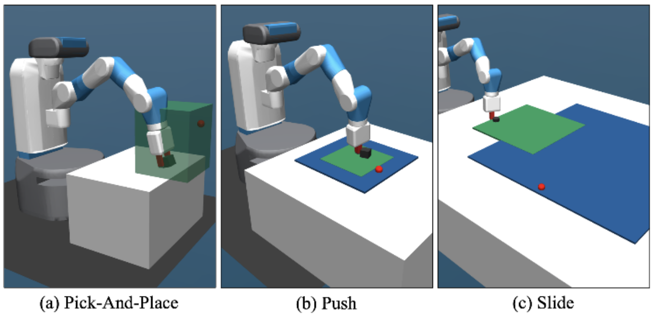

Figure 1 Overview of the benchmark tasks.

Supervisor: Prof. Alois Knoll

Advisor: Erdi Sayar, Zhenshan Bing

erdi.sayar@tum.de bing@in.tum.de

Lehrstuhl für Echtzeitsysteme und Robotik,

Fakultät für Informatik, Technische Universität München

[1] Andrychowicz, Marcin, et al. "Hindsight experience replay." Advances in Neural Information Processing Systems. 2017.
[2] Ren, Zhizhou, et al. "Exploration via Hindsight Goal Generation." NeurIPS (2019).
[3] Rui Zhao, Volker Tresp „Energy-Based Hindsight Experience Prioritization
[4] Rui Zhao, Xudong Sun, and Volker Tresp. Maximum entropy-regularized multi-goal reinforcement learning.
[5] N.Vulin et al. "Improved Learning of Robot Manipulation Tasks Via Tactile Intrinsic Motivation"