Multi-environment RL Agent for Robotics Tasks

Panagiotis Argyrakis, Revant Mahajan

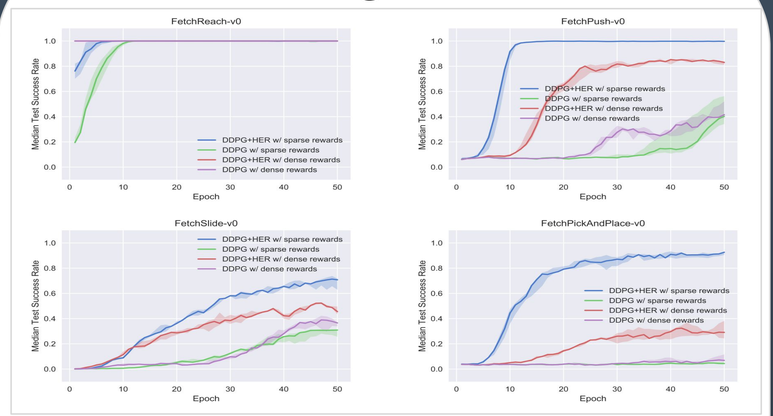
Advisors: Yanhua Li

Motivation

In this project, we aimed to explore the field of training generalized reinforcement learning models. Rather than training a new model for each different task at hand, a generalized model would create a policy that performs well across several environments. Such an agent would significantly improve the capabilities and range of tasks robots can perform.

Problem Statement:
How do we train a reinforcement learning agent on multiple different environments at once?

Background



Environment: Simulated robotic arms released by OpenAI. The arms have 7 DOF and a two-fingered parallel gripper. We used the following two:

1. FetchPush: Push a box to the desired goal position. The robot fingers are locked to prevent grasping.
2. FetchPickAndPlace: Move a box to a position in 3D space, including up in the air.

Reward Structure: Sparse Binary

Observation Structure:

1. The absolute position of the gripper
2. The position of an object with respect to the gripper

Techniques Used:

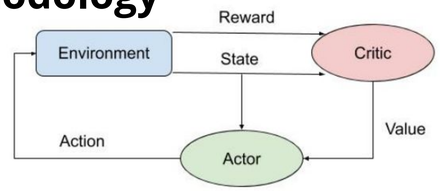
DDPG + Hindsight Experience Replay

OpenAI Benchmark Results (metric: Success Rate):

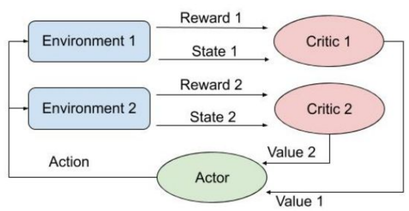
FetchPush: **1.0** FetchPickAndPlace: **0.9**

Methodology

1. **Baseline**
   Uses just one critic to evaluate both environments and to update the policy



2. **Dual Critic Approach**
   Uses two critics to evaluate both the environments separately and using the value to update the common policy

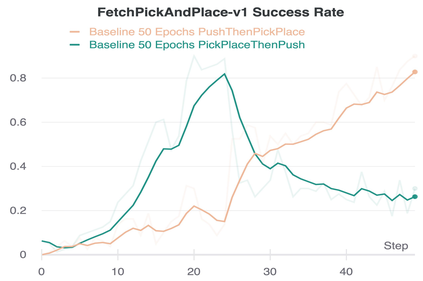


3. **Environment Aware Approach**
   Observations modified to make the model aware of which environment was being processed by injecting an extra parameter into the observation.

Experiments

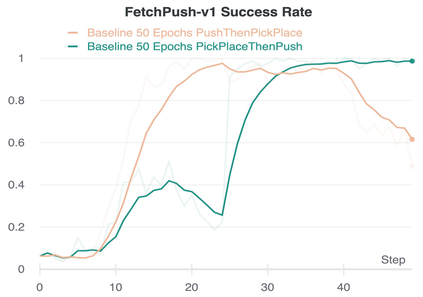
PushThenPickPlace, PickPlaceThenPush, Epoch Interlaced, Cycle Interlaced

Results - Sequential Training



Green Lines:
PickPlaceThenPush

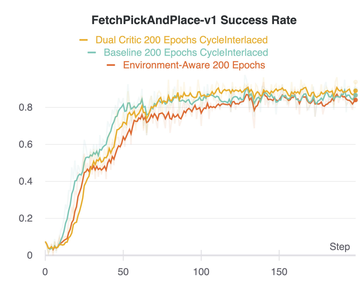
The score only increases for the environment being trained. When the environment switch occurs at epoch 25, both lines sharply change.

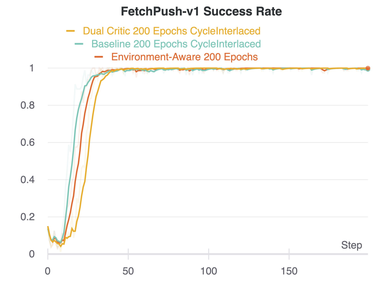


Orange Lines:
PushThenPickPlace

The score for both environments increases until epoch 40. This indicates that a major portion of FetchPickAndPlace can be achieved by FetchPush techniques.

Results - Interlaced Training





These baseline, dual-critic and environment-aware agents were trained by switching the underlying environment at every cycle, i.e. 50 times per epoch, for 200 epochs.

Using this technique, all agents achieved scores similar to the benchmarks published by OpenAI despite having to act on 2 different environments. The dual-critic agent achieved a score at FetchPickAndPlace of 0.93, 3% above the CycleInterlaced baseline.

Conclusion

- The dual-environment agents matched the performance of the OpenAI single-environment benchmark agents
- The Cycle Interlaced training technique produced the best performing agents and the lowest variance during training
- The baseline agents performed comparably to the dual-critic and environment-aware agents
- Any FetchPickAndPlace task with a target location on the platform can be performed as a FetchPush task.

Resources

GitHub: github.com/panargirakis/hindsight-experience-replay
Weights & Biases:
wandb.ai/pargyrakis/Input-Based%20DDPG%20+%20HER%20Dual%20Fetch%20Model

Acknowledgements

TianhongDai for the simplified implementation of DDPG + HER for the Fetch robotics environments