Combining Hindsight Experience Replay with Hierarchical Reinforcement Learning

Armaan Sethi

I. Introduction

THERE has been a wide range of success in learning policies for sequential decision-making tasks by using reinforcement learning combined with neural networks. These successes range from perfecting Atari games, defeating the best human Go player, hitting a baseball, screwing a cap on a bottle, and door opening to helicopter control. One difficulty for reinforcement learning is to learn with very sparse rewards, however, recently OpenAI has created an algorithm that works exceptionally well for these tasks called Hindsight Experience Replay (HER). HER is an algorithm for reinforcement learning that specializes in learning from failure. When a failure occurs, instead of just recording the actions taken lead to a failure, it records that the actions could have led to a success if the goal was in a different place. Another common approach in reinforcement learning is using Hierarchical Reinforcement Learning (HRL) in which a higher-level policy generates subgoals that should be learned. It is possible to combine HER and HRL. Instead of applying HER just to goals, it could also be applied to actions generated by the higher-level policy.

II. Related Work

The idea for this project came from a paper from OpenAI that introduced a suite of challenging continuous control tasks and presented a set of concrete research ideas for improving reinforcement learning algorithms. This paper released environments for existing robotic hardware, such as the Fetch robotic arm and the Shadow Dexterous Hand. I only used the Fetch environments, specifically on the ”Reaching”, ”Pushing”, ”Sliding”, and ”Pick & Place” tasks. I will go more in depth on these environment in section 3. This project was mostly based upon two recent papers improving different aspects of reinforcement learning.

A. Hindsight Experience Replay

One of the biggest challenges in Reinforcement Learning is dealing with sparse rewards. Hindsight Experience Replay is a method used in order to learn efficiently from sparse binary rewards. This avoids the need for complicated reward engineering. Hindsight Experience Replay specializes in learning from failure. When a failure occurs, instead of just recording the actions taken lead to a failure, it records that the actions could have led to a success if the goal was in a different place.

B. Hierarchical Reinforcement Learning

Hierarchical Reinforcement learning is an approach used to tackle complicated problems in which the Reinforcement Learning process is divided into a hierarchy of actions at different levels of abstraction. By decomposing a large problem into many smaller ones it is easier to transfer knowledge from one task to another and to learn complicated tasks.

III. Problem Definition

The problems used are defined by the environments provided by OpenAI. The tasks including reaching, pushing, sliding, and a more complicated ”pick & place” task. These tasks are based on the 7-DoF Fetch robotics arm, which has a two-fingered parallel gripper. In the reaching task, the Fetch robot must move the gripper to a specified target position. In the pushing task, the Fetch robot must move a box, that is placed on a table, to a target location also on the table. In the sliding task, a puck is placed on a long slippery table and the target is out of the robot’s reach. This forces the robot to hit the puck with a force in order for the puck to stop at the location due to friction. In the pick & task task, the Fetch robot must grasp a vox and move it to the target location that may be located on the table surface or in the air above it.

IV. Methods

A. Implemented

I have read the Hindsight Experience Replay Paper by OpenAI. I have read a couple papers describing various implementations of Hierarchical Reinforcement Learning. I have implemented Hindsight Experience Replay for the four environments provided by OpenAI without the hierarchical approach. I have split the tasks into various subtasks. These tasks were already relatively simple, so there were very few subtasks per task - not sure how this will affect things.

B. Implementing

I plan to solve this problem by applying Hindsight Experience Replay algorithm to these subtasks.

V. Results

I hope to see the success rate of the tasks to improve quicker when combining HER and HRL, than just HER alone. I also hope to show that subtasks allow transfer of knowledge.

VI. Conclusion and Future Work

4/17: Implement and test my method on the reaching environment.
4/19: Implement and test method on the other environments.
4/21: Finish results
4/23: Finish writing

Acknowledgment

Thanks to Professor Alterovitz for teaching and giving this opportunity.

References

[1] Matthias Plappert, Marcin Andrychowicz, Alex Ray, Bob McGrew, Bowen Baker, Glenn Powell, Jonas Schneider, Josh Tobin, Maciek Chociej, Peter Welinder, Vikash Kumar, Wojciech Zaremba. (2018). Multi-Goal Reinforcement Learning: Challenging Robotics Environments and Request for Research. *arXiv preprint arXiv:1802.09464.*

[2] Levy, A., Platt, R., and Saenko, K. (2017). Hierarchical actor-critic. *arXiv preprint arXiv:1712.00948.*

[3] Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Abbeel, O. P., and Zaremba, W. (2017). Hindsight experience replay. In Advances in *Neural Information Processing Systems*, pages 50555065.

[4] Dayan, Peter and Hinton, Geoffrey E. Feudal reinforcement learning. In Hanson, S. J., Cowan, J. D., and Giles, C. L. (eds.), Advances in *Neural Information Processing Systems 5*, pp. 271278. Morgan-Kaufmann, 1993.