

CS 330 Autumn 2021/2022 Homework 3

Goal Conditioned Reinforcement Learning and Hindsight Experience Replay

Due Wednesday October 27th, 11:59 PM PST

SUNetID: tminh
Name: Minh Tran
Collaborators: N/A

October 23, 2021

Abstract

The document contains solutions for adapting existing Deep Q-Network to be goal-conditioned and the implementation of Hindsight Experience Replay on top of a goal-conditioned DQN. This also includes experiments and explanations on Bit Flipping and Sawyer Reach environments.

1 Implementing Goal-conditioned RL

Details of implementation is in `run_episode.py`. At every step, the state vector and the goal vector were concatenated and passed to model to get output action. The episode experience, episodic reward, and state were updated accordingly after the decision was made.

2 Adding HER to Bit Flipping

Implementation of HER is in file `trainer.py`, function `update_replay_buffer`. At each episode in the experience, label of episode was relabeled according to the argument, either future, final or random and the reward was also recalculated. The new concatenated data then was passed to the replay buffer to enrich the data for training process.

3 Analyzing HER for Bit Flipping Environment

a. 6 bits 250 epochs

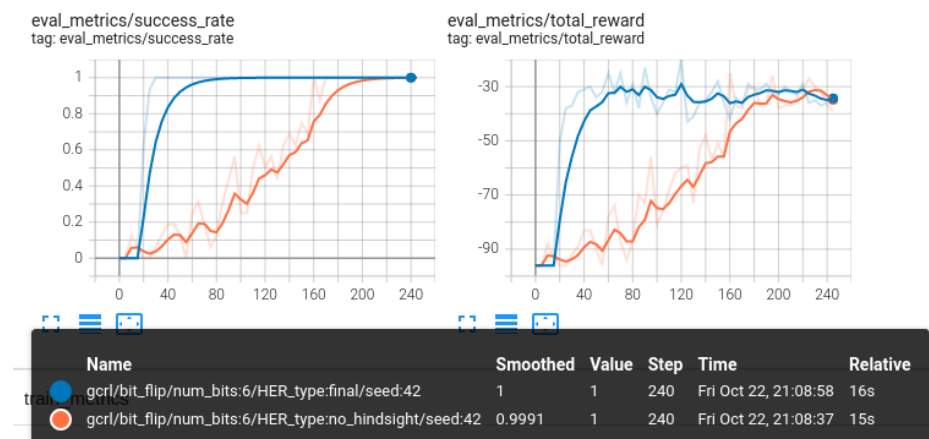


Figure 1: Bit Flipping performance with and without HER (blue/orange)

b. 15 bits 500 epochs

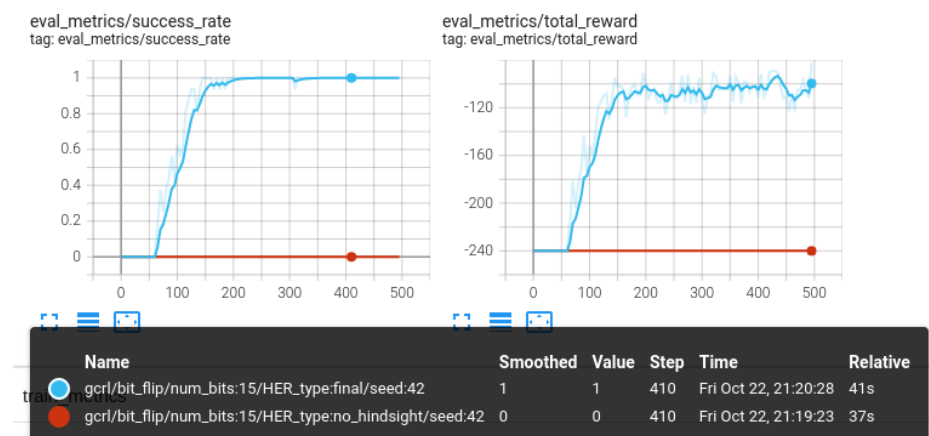


Figure 2: Bit Flipping performance with and without HER (blue/red)

c. 25 bits 1000 epochs

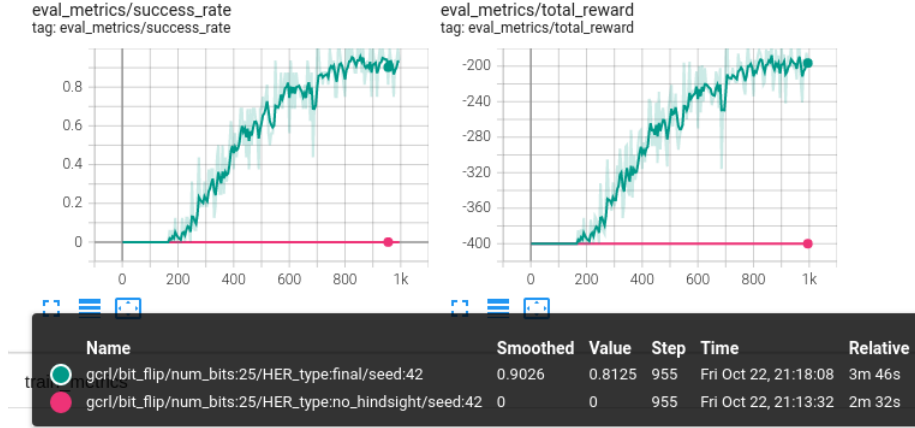


Figure 3: Bit Flipping performance with and without HER (green/red)

d. 15 bits 500 epochs

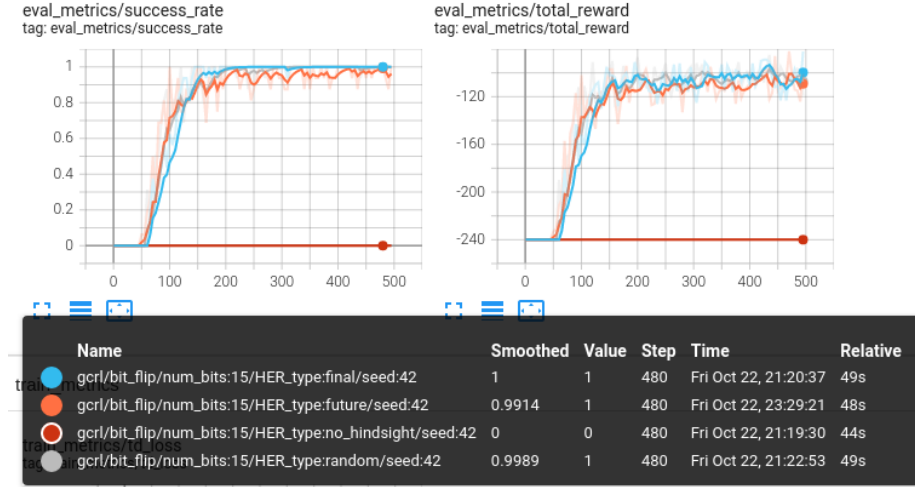


Figure 4: Bit Flipping performance with and without HER (others/red)

e. In general, we observed that with the number of bit increases, the complexity of the task becomes more challenging for our based model since there are too many impact factors while the data is not enough, it hardly finished on the first setting while failed for larger number of bits.

On the other hand, HER performed really well no matter the method, which suggests additional data from early training is really useful for complex task with limited data.

4 Analyzing HER for Sawyer Reach

a. Tensorboard

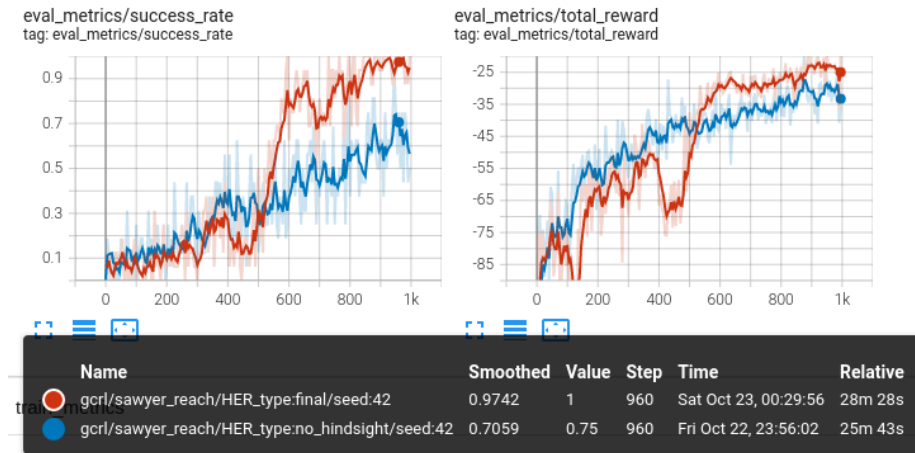


Figure 5: Sawyer Reach performance with and without HER (red/blue)

b. Discussion

From the Figure 5, we can see that for Sawyer Reach, HER significantly helped in improving performance of the task. while at the beginning, the two models behaved pretty similar, but around 500 steps, model with HER improved a lot, I think that because at that point, we collected a good amount of data that is enough to help model to learn. The normal have a stable success increase per episode instead of a tipping point like HER.

Meanwhile, from figures in previous section, we can see that for Bit Flipping, as the number of bit increases, the complexity of the task increases so it is harder to learn the task, so without HER, the model performed really poor while any other HER method improved the agent significantly.