CS 330 Autumn 2021/2022 Homework 3: Goal Conditioned Reinforcement Learning and Hindsight Experience Replay

Kevin Lee (kelelee@stanford.edu)

Due on Wednesday October 27th, 11:59 PM PT

3 Analyzing HER for Bit Flipping Environment

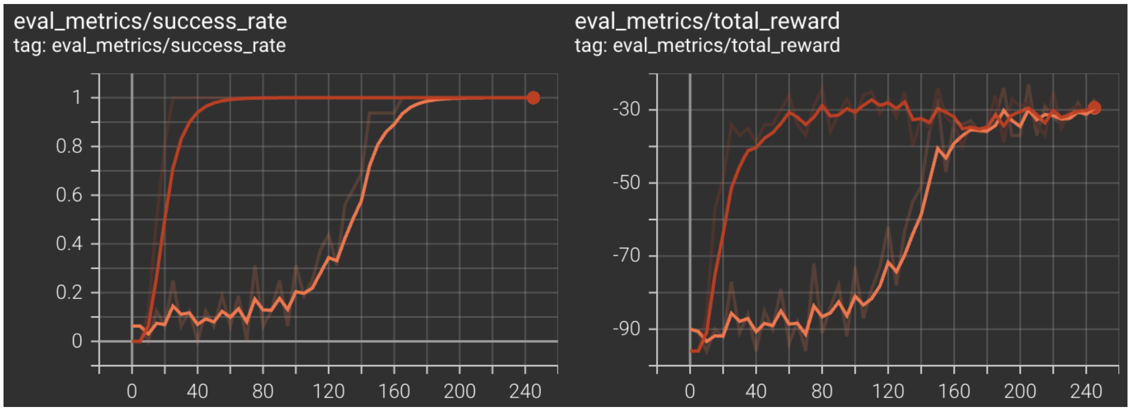

Figure 1: Eval metrics for bit flip environment with 6 bits, without HER (orange) and with final-type HER (red)

Answer:

As expected, the attempt with episodes augmented with HER would reach a 100% success rate earlier than the attempt without HER. This is because there would be more (s, a, r, s', g) tuples where the rewards are positive, allowing the model to learn how to select the right action, a given the goal, g much better.

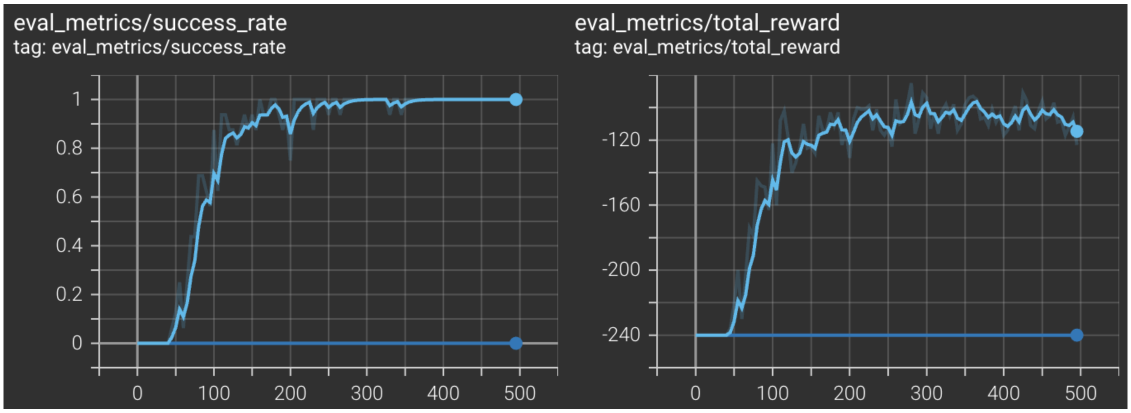

Figure 2: Eval metrics for bit flip environment with 15 bits, without HER (dark blue) and with final-type HER (light blue)

Answer:

As in the previous case, the attempt with HER has more success in reaching the desired goal. However, due to the bigger action and state space, the model without HER was not able to encounter the goal state within 500 attempts, hence the 0 success rate throughout the entire training. On the other hand, do also note that the attempt with HER took longer than the previous case (6-bit) which is expected due to the larger action and state space.

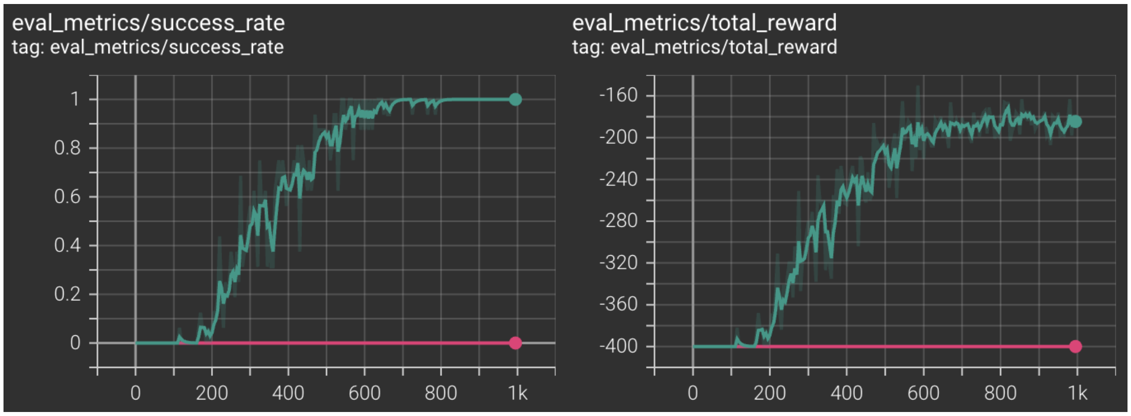

Figure 3: Eval metrics for bit flip environment with 25 bits, without HER (pink) and with final-type HER (green)

Answer:

Very similar to the previous case, as expected. Due to the bigger action and state space, the attempt without HER remained at 0 success rate while the attempt with HER took slightly longer than in the 15-bit case.

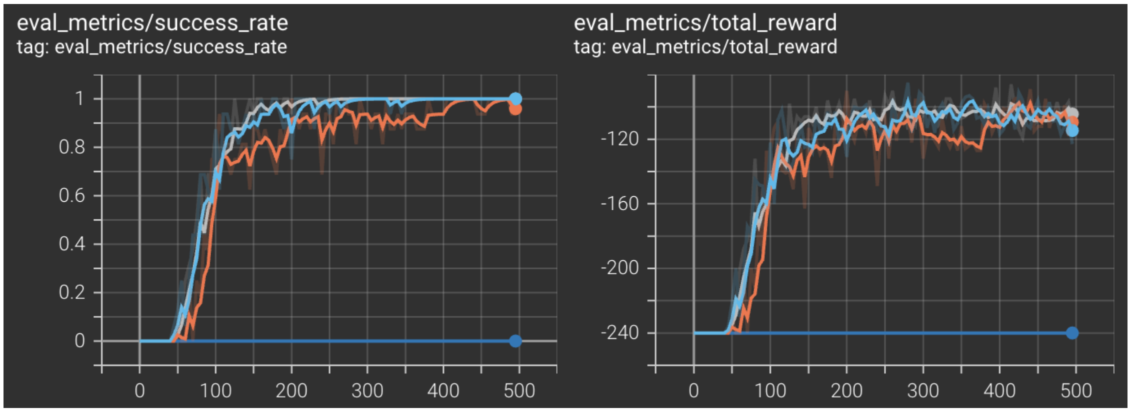

Figure 4: Eval metrics for bit flip environment with 15 bits, without HER (dark blue), with final-type HER (light blue), random-type HER (white) and future-type HER (orange)

Answer:

All three types of HER seem to provide very similar quality of data augmentation, helping the model converge to a good policy within 200 steps. One might expect *future*-type HER to perform better than *final* and *random* because *future*-type HER seems to augment the episodes in a more relevant way, but in the bit flip setting, the goal could be any possible combination of bits and hence any (s, a, r, s', g) episode would be just as valuable.

4 Analyzing HER for Sawyer Reach

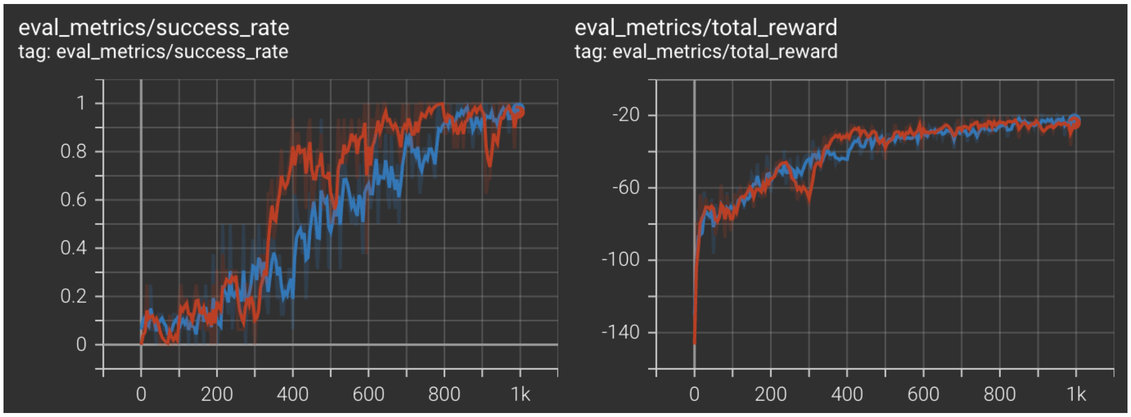

Figure 5: Eval metrics for Sawyer reach environment, without HER (dark blue) and with final-type HER (red)

Answer:

While HER still helps the model reach high success rate earlier in Sawyer reach, the improvement isn't as drastic as in the bit flip environment. This is because the reward in the Sawyer environment is $r = -||s - s'||^2$, which means that even without HER, the reward in state space is not sparse, allowing the model to infer good transitions from bad ones.