

TME 12: # Implicit Curriculum RL

In this TME, we explore the methods of Implicit Curriculum RL for sparse-rewards environment.

DQN with Goal

Our first environment `plan2Multi` has a method `sampleGoal` that uniformly samples an intermediate state-goal among a set of predefined goals. With that in mind, we can reuse a simple Double DQN agent: instead of learning a mapping from state-action pairs to values, the agent now learns a mapping from an 3-uplet state, action, goal. In practice, we simply concatenate the current episode's goal to extend the observation. We plot the test reward, computed on 100 episodes every 100 iterations:

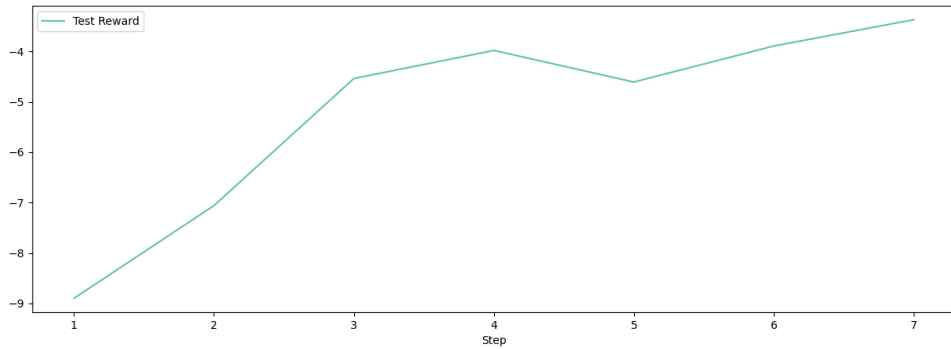
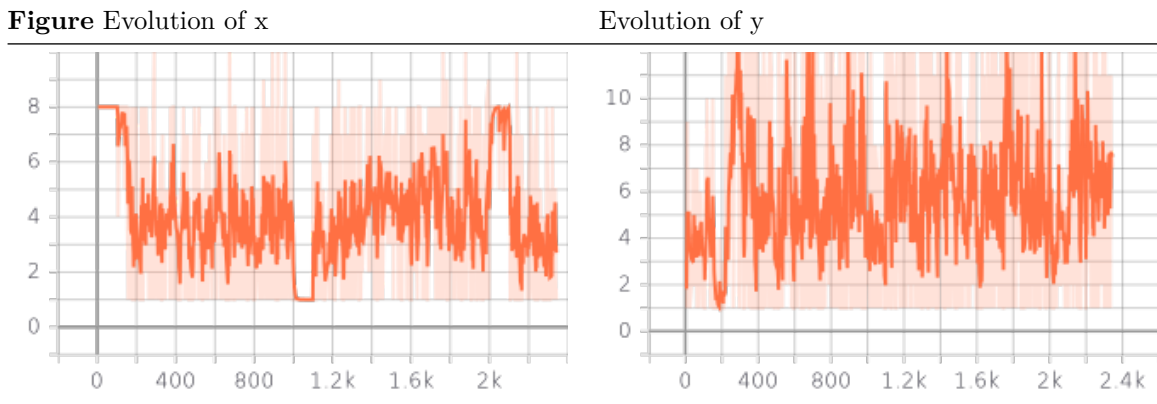


Figure 1: add image

Our second environment `plan2` does not come with a set of predefined goals to sample from. With only one goal, relatively far away from its starting state, the agent never learns to navigate in the grid away from its initial state, as we can observe by looking at (x,y) , the final coordinated of the agent at the end of each episode:



Hinsight Experience Replay

We implement the Hinsight Experience Replay (HER), which fills the Replay Buffer of our agent with new transitions, with the last visited state at the end of episode as the transition's goal. The HER is implemented as a wrapper to the agent. At each call of the `act` function, it checks if the episode is done, creates the supplementary transitions and adds them in the agent's buffer.

As a result, our cumulative rewards starts to increase at the third testing session, as we observe the agent terminating more and more frequently its episode in the final stat: (33,38):

