

Hindsight Experience Replay

esgl Hu

March 27, 2018

Algorithm 1 Hindsight Experience Replay (HER)

Given:

- an off-policy RL algorithm \mathbb{A} , ▷ e.g. DQN, DDPG, NAF, SDQN
- a strategy \mathbb{S} for sampling goals for replay, ▷ e.g. $\mathbb{S}(s_0, \dots, s_T) = m(s_T)$
- a reward function $r : \mathcal{S} \times \mathcal{A} \times \mathcal{G} \rightarrow \mathbb{R}$. ▷ e.g. $r(s, a, g) = -[f_g(s) = 0]$

Initialize \mathbb{A}

▷ e.g. initialize neural networks

Initialize replay buffer R

for episode=1, M **do**

 Sample a goal g and an initial state s_0 .

for $t = 0, T - 1$ **do**

 Sample an action a_t using the behavioral policy from \mathbb{A} :

$a_t \leftarrow \pi_b(s_t || g)$

▷ $||$ denotes concatenation

 Execute the action a_t and observe a new state s_{t+1}

end for

for $t = 0, T-1$ **do**

$r_t := r(s_t, a_t, g)$

 Store the transition $(s_t || g, a_t, r_t, s_{t+1} || g)$ in R

▷ standard experience replay

 Sample a set of additional goals for replay $G := \mathbb{S}(\text{current episode})$

for $g' \in G$ **do**

$r' := r(s_t, a_t, g')$

 Store the transition $(s_t || g', a_t, r', s_{t+1} || g')$ in R

end for

end for

for $r = 1, N$ **do**

 Sample a minibatch B from the replay buffer R

 Perform one step of optimization using \mathbb{A} and minibatch B

end for

end for
