

Deep reinforcement learning with robust deep deterministic policy gradient

ABSTRACT

Recently, Deep Deterministic Policy Gradient (DDPG) is a popular deep reinforcement learning algorithms applied to continuous control problems like autonomous driving and robotics. Although DDPG can produce very good results, it has its drawbacks. DDPG can become unstable and heavily dependent on searching the correct hyperparameters for the current task. DDPG algorithm risk overestimating the Q values in the critic (value) network. The accumulation of estimation errors as time elapse can result in the reinforcement agent trapping into a local optimum or suffering from disastrous forgetting. Twin Delayed DDPG (TD3) mitigated the overestimation bias problem but might not exploit full performance due to underestimation bias. In this paper Twin Average Delayed DDPG (TAD3) is proposed for specific adaption to TD3 and shows that the resulting algorithm perform better than TD3 in a challenging continuous control environment.