INTERPRETABLE META-REINFORCEMENT LEARNING WITH ACTOR-CRITIC METHOD

Anonymous authors
Paper under double-blind review

ABSTRACT

Meta-reinforcement learning (meta-RL) algorithms have successfully trained agent systems to perform well on different tasks within only few updates. However, in gradient-based meta-RL algorithms, the Q-function at adaptation step is mainly estimated by the return of few trajectories, which can lead to high variance in Q-value and biased meta-gradient estimation, and the adaptation uses a large number of batched trajectories. To address these challenges, we propose a new meta-RL algorithm that can reduce the variance and bias of the meta-gradient estimation and perform few-shot task data sampling, which makes the meta-policy more interpretable. We reformulate the meta-RL objective, and introduce contextual Q-function as a meta-policy critic during task adaptation step and learn the Q-function under a soft actor-critic (SAC) framework. The experimental results on 2D navigation task and meta-RL benchmarks show that our approach can learn an more interpretable meta-policy to explore unknown environment and the performance are comparable to previous gradient-based algorithms.