Summary of Proximal Policy Optimization

Thursday, December 31, 2020 10:46 PM

PPO Summary:

1. Collect trajectories based on policy π_θ, and initialize $\theta' = \theta$

2. Next, Compute the gradient of the cliped surrogate function using trajectories.

3. Update θ' using gradient ascent
$\theta' \leftarrow \theta' + \alpha \nabla_{\theta'} L^{clip}_{sur}(\theta', \theta)$

4. Repeat step 2-3 without generating new trajecties. Only a few times of repeating is allowed.

5. Set $\theta = \theta'$, go back to Step 1, repeat!