Proximal Policy Evolution

April 27, 2020

PPO Surrogate Loss Objective,

$$L^{CLIP}(\theta) = \mathbf{E}_t[min(r_t(\theta)\tilde{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\tilde{A}_t)]$$

And the ES update-

$$\theta \leftarrow \theta + \frac{\alpha}{n\sigma}\sum_{i=1}^{n} F_i\epsilon_i$$

Combining the update with the objective gives us Proximal Policy Evolution. The algorithm augments a policy $\pi(a|s;\theta)$ by evolving its parameters θ and optimizing them using the clipped objective. All notations are followed as per convention and k is the number of winners in a population.

Algorithm 1 Proximal Policy Evolution

Initialize k, θ
for all $t = 0, 1, 2..$ **do**
    sample $\epsilon_1, \epsilon_2, ...\epsilon_n \sim \mathcal{N}(0, 1)$
    compute returns $\mathrm{F}_i = F(\theta + \sigma\epsilon_i)$ for $i = 1, 2, ...n$
    sort F in decreasing order
    set $\theta_t \leftarrow \theta_t + \frac{\alpha}{n\sigma}\sum_{i=1}^{n} F_i\epsilon_i$ where $F_i \in F_1, F_2, ..., F_k$
    run π_{θ_t} in the env
    compute $\tilde{A}_t$
    $L^{CLIP}(\theta) \leftarrow \mathbf{E}_t[min(r_t(\theta)\tilde{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\tilde{A}_t)]$ for j epochs
    $\theta_{t+1} \leftarrow \theta_t$
end for