Reviewing Natural Actor Critic methods

Maximilian Gehrke
Tabea Wilke
Yannik Frisch

Group 19 Oleg Arenz

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Natural Gradient

TECHNISCHE UNIVERSITÄT DARMSTADT

- Optimization problem:

$$\max_{\delta\theta} J(\theta + \delta\theta) \approx J(\theta) + \delta\theta^T \nabla_\theta J(\theta)$$
$$\text{s.t. } \epsilon = D_{KL}(\pi_\theta || \pi_{\theta+\delta\theta}) \approx \tfrac{1}{2}\delta\theta^T F_\theta \delta\theta$$

- Solution:

$$\widetilde{\nabla}_\theta J(\theta) = F_\theta^{-1} \nabla_\theta J(\theta).$$

- Fisher Information Matrix:

$$F_\theta = \mathbb{E}_{\pi_\theta}\left[\nabla_\theta \log \pi_\theta(a|s) \nabla_\theta \log \pi_\theta(a|s)^T\right]$$

The Natural Actor Critic algorithm

TECHNISCHE UNIVERSITÄT DARMSTADT

- Compatible Function Approximation:

$$\nabla_w A_w(s,a) = \nabla_\theta \log \pi_\theta(a|s)$$
$$A_w(s,a) = \nabla_\theta \log \pi_\theta(a|s) w$$

- Actor update:

$$\begin{aligned}\nabla_\theta J(\theta) &= \mathbb{E}_{\pi_\theta}\left[\nabla_\theta \log \pi_\theta(a|s) A_w(s,a)\right] \\ &= \mathbb{E}_{\pi_\theta}\left[\nabla_\theta \log \pi_\theta(a|s) \nabla_\theta \log \pi_\theta(a|s)^T w\right] = F_\theta w \\ \widetilde{\nabla}_\theta J(\theta) &= w\end{aligned}$$

- Critic update: Episodic $\Leftrightarrow$ LSTD(λ)

Extensions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Recursive Least Squares
- Fitted NAC + Importance Sampling
- Incremental NAC
- Implicit Incremental NAC

Conclusion & Discussion

TECHNISCHE UNIVERSITÄT DARMSTADT