

Deep Deterministic Policy Gradient (DDPG) for Atari Galaga

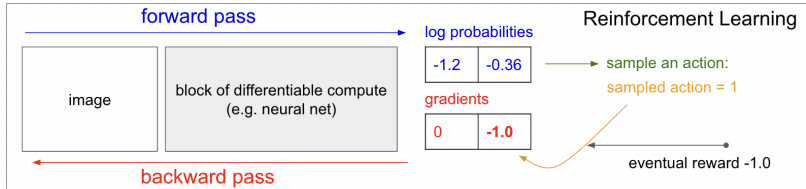
CS4100 Project Proposal

Thomas Wilson

Khoury College of Computer Sciences

March 10, 2022

Policy Gradient Methods



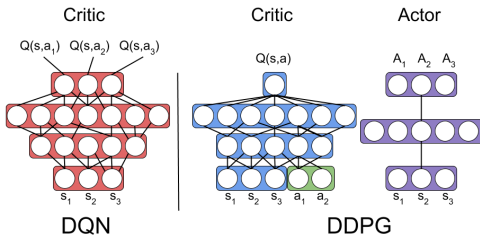
(note that this uses stochastic policy)

Policy Gradient Methods

$$\begin{aligned} J(\mu_\theta) &= \int_{\mathcal{S}} \rho^\mu(s) r(s, \mu_\theta(s)) ds \\ &= \mathbb{E}_{s \sim \rho^\mu} [r(s, \mu_\theta(s))] \end{aligned} \quad (8)$$

$$\begin{aligned} \nabla_\theta J(\mu_\theta) &= \int_{\mathcal{S}} \rho^\mu(s) \nabla_\theta \mu_\theta(s) \nabla_a Q^\mu(s, a)|_{a=\mu_\theta(s)} ds \\ &= \mathbb{E}_{s \sim \rho^\mu} \left[\nabla_\theta \mu_\theta(s) \nabla_a Q^\mu(s, a)|_{a=\mu_\theta(s)} \right] \end{aligned} \quad (9)$$

Actor-Critic



- 2 Networks:
critic: $(s, a) \rightarrow V(s')$
actor: $s \rightarrow a$
- Continuous Actions
- Action-Replay Buffer
- Target Network Updates

Frameworks

- OpenAI Gym
- PyTorch/Autograd
- OpenCV
- LibVirt

Approach

- 1 Set up environment
- 2 Implement Deep *Stochastic* Policy Gradient Algorithm
- 3 Add Replay Buffer, Target Network
- 4 Implement Actor-Critic For Deterministic Policy Gradient Algorithm

References

- DeepMind's 'Continuous Control with Deep Reinforcement Learning' (2016)
- DeepMind's 'Deterministic Policy Gradient Algorithms' (2014)
- DeepMind's 'Playing Atari with Deep Reinforcement Learning' (2015)
- Karpathy's 'Deep Reinforcement Learning: Pong from Pixels' (<https://karpathy.github.io/2016/05/31/rl/>)