Asymmetric DQN for Partially Observable Reinforcement Learning

UAI 2022, Eindhoven, Netherlands

Andrea Baisero Brett Daley Christopher Amato
{baisero.a,daley.br,c.amato}@northeastern.edu
Northeastern University, Boston, USA

Asymmetric DQN for Partially Observable Reinforcement Learning

Setting:

- Single-agent partially observable reinforcement learning.
- Offline training in simulation $\implies$ exploit state for training.

History-State Values $U^\pi(h, s, a)$**:**

$$U^\pi(h, s, a) = R(s, a) + \gamma \, \mathbb{E}_{s', o|s,a} \left[U^\pi(hao, s', \pi(hao)) \right]$$

Methodology: Train $\hat{U}(h, s, a)$, $\hat{Q}(h, a)$ jointly.

Asymmetric DQN:

$$\mathcal{L}_{\hat{U}} = \left(r + \gamma \hat{U}(hao, s', \operatorname*{argmax}_{a'} \hat{Q}(hao, a')) - \hat{U}(h, s, a) \right)^2$$

$$\mathcal{L}_{\hat{Q}} = \left(r + \gamma \hat{U}(hao, s', \operatorname*{argmax}_{a'} \hat{Q}(hao, a')) - \hat{Q}(h, a) \right)^2$$

Northeastern University