Model-free Deep Reinforcement Learning (DRL) Algorithms 无模型深度强化学习算法总结

Dynamic Programming and Markov Decision Processes

Actor-Critic Methods. 2000

stochastic policy

Natural Policy Gradient. 2002

trust region

TRPO. Trust Region Policy Optimization. 2015

GAE. Generalized Advantage Estimation. 2015

advantage function

A3C. Asynchronous A2C. 2016

synchronous

A2C. Advantage Actor-Critic. 2016

trust region (approximated) KL penalty coefficient

PPO. Proximal Policy Optimization. 2017

Energy-Based Policy

SQL. Soft Q-learning. 2017

maximum entropy automatic temperature tuning

SAC. Soft Actor-Critic. 2018

Q-learning. 1992

Q-table $\rightarrow Q$ net. experience replay

DQN. Deep Q Network. 2014

Q net. \rightarrow 2 Q net.

Double DQN. 2016

deterministic policy

DPG. Deterministic Policy Gradient. 2015

Taming the Noise via soft update. 2015

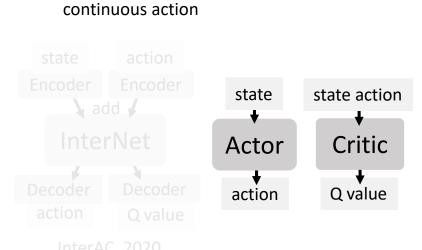
greedy policy \rightarrow policy net.

DDPG. Deep DPG. 2016

2 Q net. → Twin Critic delay target update policy smoothing (SPG in DPG)

TD3. Twin Delayed DDPG. 2018

state Q net $arg max(Q_1, Q_2, . Q_n)$ action discrete action



3 Python file for Deep RL in PyTorch (lightweight) https://github.com/Yonv1943/DL RL Zoo See more in https://lilianweng.github.io/lil-log/2018/04/08/policy-gradient-algorithms.html 中文翻译见 https://tomaxent.com/2019/04/14/策略梯度方法/ 译者为 Ewan Li 此图制作于 2020-04-03