Deep Reinforcement Learning (DRL) Algorithms 深度强化学习算法总结 Dynamic Programming and Markov Decision Processes (MDPs)

VPG. Vanilla Policy Gradient. 2000 AC. Actor-Critic Methods. 2000

Stochastic Policy

♦ NPG. Natural Policy Gradient. 2002

Trust region

TRPO. Trust Region Policy Optimization. 2015

GAE. Generalized Advantage Estimation. 2015

Advantage Function Advantage Actor-Critic

A3C. Asynchronous A2C. 2016

Trust Region (approximated, surrogate) KL penalty coefficient

PPO. Proximal Policy Optimization. 2017

Auxiliary Task (on-policy, off-policy)

PPG. Proximal Policy Gradient. 2020

Energy-Based Policy

SQL. Soft Q-learning. 2017

maximum entropy automating entropy adjustment

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

+Advantage function

Dueling 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

Distributional Perspective

C51 DQN (Categorical 51 grids). 2017

Quantile Regression

QR-DQN. 2017

Prioritized sweeping. 1993

PER. Prioritized Experience Replay. 2016 HER. Hindsight Experience Replay. 2017

All DON Variances

Rainbow DON, 2017

D4PG(Distributed Distributional DDPG). 2017

Ape-X DQN. 2018

Ape-X DPG. 2018

Use Monte Carlo Tree Search 2006

AlphaGo. 2015

Without human knowledge

AlphaZero. 2017

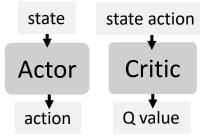
Value Prediction Network. 2017

Atari game 2600

MuZero. 2019 (Model-based RL)

off-policy on-policy state Q net arg max($Q_1, ..., Q_n$) action discrete action space

continuous action space



parameterized action space

Parameterized-DQN. 2018 output a probability

SAC for Discrete Action Space. 2019

Combination of DQN and DDPG

Hybrid-PPO. 2019

强化学习小雅 Light weight and Elegant PyTorch DRL in https://github.com/Yonv1943/ElegantRL 强化学习公式 https://lilianweng.github.io/lil-log/2018/04/08/policy-gradient-algorithms.html 强化学习翻译 https://tomaxent.com/2019/04/14/策略梯度方法/ 译者为 Ewan Li 此图制作时间 2021-01-12