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
A2C (Advantage Actor-Critic)
A3C. Asynchronous A2C. 2016

Trust Region (approximated, surrogate)
KL Penalty Coefficient
PPO. Proximal Policy Optimization. 2017

Planning (Model-based RL)
MBPO. Model Based PO. 2019
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 → Q net.
Experience Replay
DQN. Deep Q Network. 2014

Q net. → 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 → Policy Net.
DDPG. Deep DPG. 2016
D4PG(Distributed Distributional DDPG). 2017

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

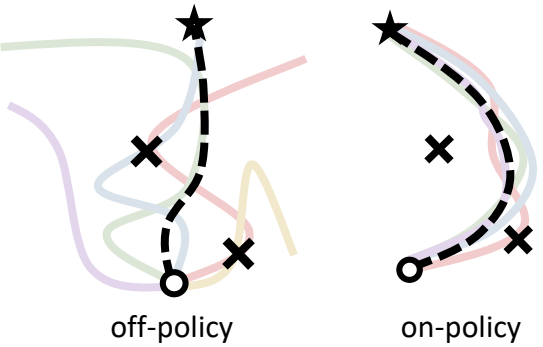
Prioritized sweeping. 1993
PER. Prioritized Experience Replay. 2016
HER. Hindsight Experience Replay. 2017

All DQN Variances
Rainbow DQN. 2017

Use Monte Carlo Tree Search 2006
AlphaGo. 2015
Without human knowledge
AlphaZero. 2017
Planning using Dynamics Model
Value Prediction Network. 2017
+Atari game (continuous state space)
MuZero. 2019 (Model-based RL)

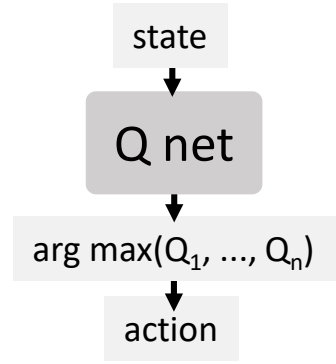
Ape-X DQN. 2018
Ape-X DPG. 2018

IMPALA. 2018
Seed RL. 2020

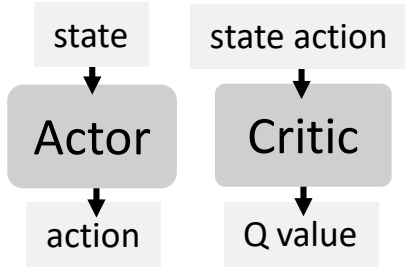


hybrid action

parameterized action space
Parameterized-DQN. 2018
output a probability
SAC for Discrete Action. 2019
Combination of DQN and DDPG
Hybrid-PPO. 2019



discrete action space
continuous action space



知乎 – **如何选择**深度强化学习算法？*MuZero/SAC/PPO/TD3/DDPG/DQN/等*
知乎 – 深度强化学习**调参技巧**：*以D3QN、TD3、PPO、SAC算法为例*