DEEP REINFORCEMENT LEARNING WORKSHOP

Alireza Nobakht • AICup

OUTLINE

Good Materials

Reinforcement Learning

Deep Reinforcement Learning

OpenAI Gym

Libraries and Frameworks for Deep RL

DQN Code Review

Q&A

GOOD MATERIALS

DeepMind Course

https://www.deepmind.com/learning-resources/reinforcement-learning-lecture-series-2021

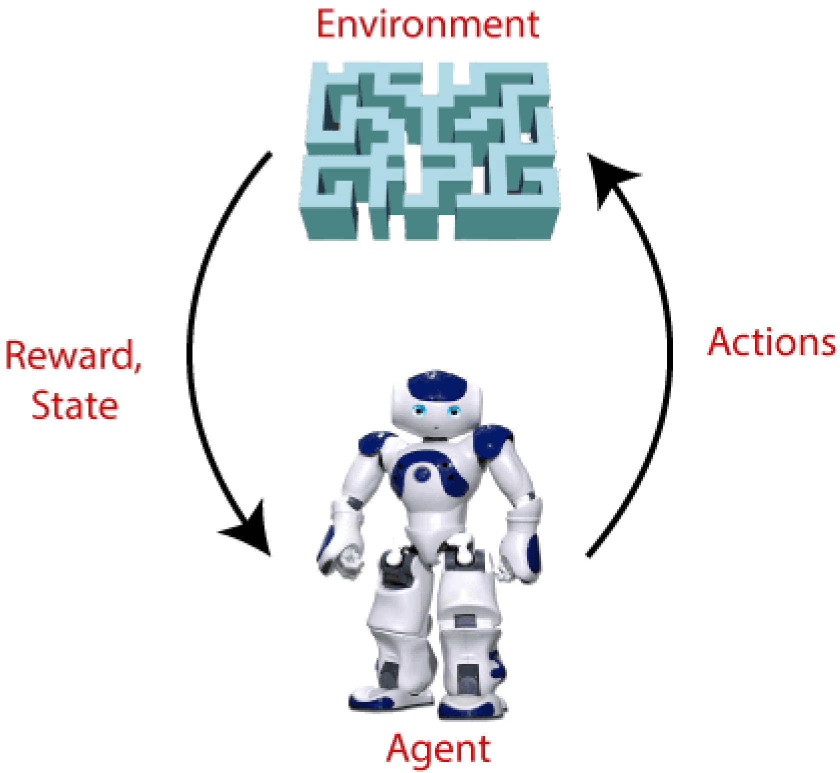
Stanford CS234

https://web.stanford.edu/class/cs234/

Reinforcement Learning: An Introduction

https://www.andrew.cmu.edu/course/10-703/textbook/BartoSutton.pdf

REINFORCEMENT LEARNING

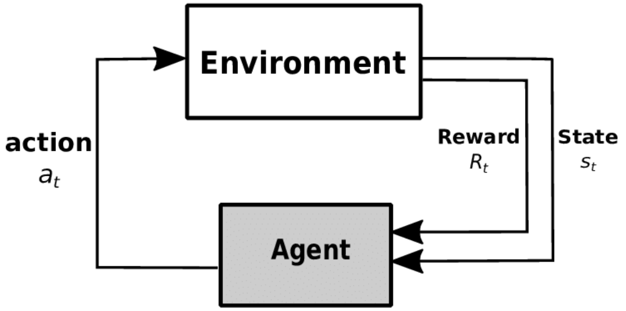


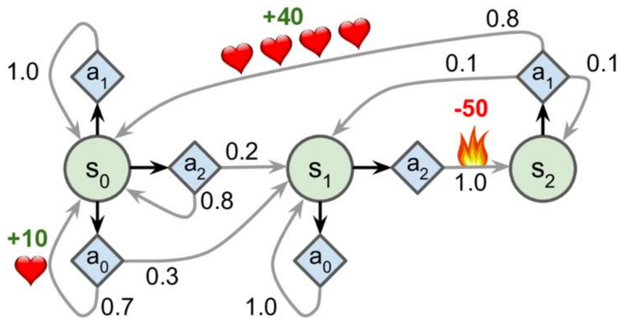
REINFORCEMENT LEARNING

MARKOV DECISION PROCESS

$$M = \langle S, A, R, T, \gamma \rangle$$

- S - state space with $s_t \in S$
- A - action space with $a_t \in A$
- R - reward function $R : S \times A \to \mathbb{R}$
- T - transition function $T(s_{t+1}|s_t, a_t)$
- γ - discount factor $0 < \gamma \leq 1$





REINFORCEMENT LEARNING

Agent selects its action based on policy function $\pi(s_t)$

Or randomly based on policy distribution function $\pi(a_t|s_t)$

The goal is to find optimal policy π^* where

$$\pi^* = \arg\max_{\pi} \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)]$$

important auxiliary functions

- Q function $Q(s_t, a_t)$
- Value function $V(s_t)$

REINFORCEMENT LEARNING

Q-LEARNING

Find optimal Q-function using bellman equation

$$Q(s,a) \leftarrow R(s,a) + \gamma \max_{a'} Q(s',a')$$

Then optimal policy is

$$\pi^*(s_t) = \arg\max_{a} Q^*(s_t,a)$$

REINFORCEMENT LEARNING

TYPES OF ALGORITHMS

Model based

$T(s_{t+1}|s_t, a_t)$

Model free

- Value based

  $Q(s_t, a_t)$

- Policy based

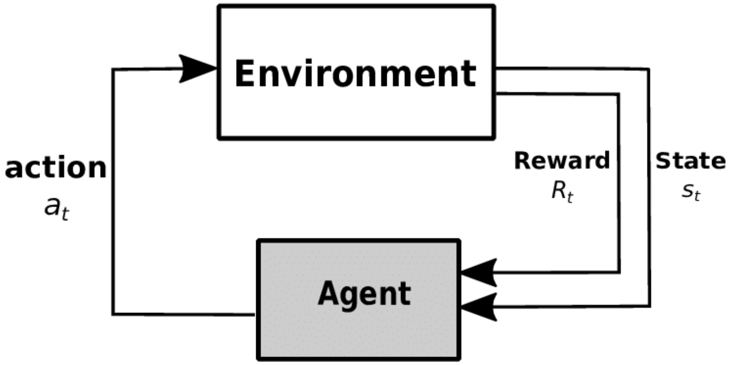
  $\pi(s_t)$

- Hybrid (actor-critic)

  $Q(s_t, a_t)$ + $\pi(s_t)$

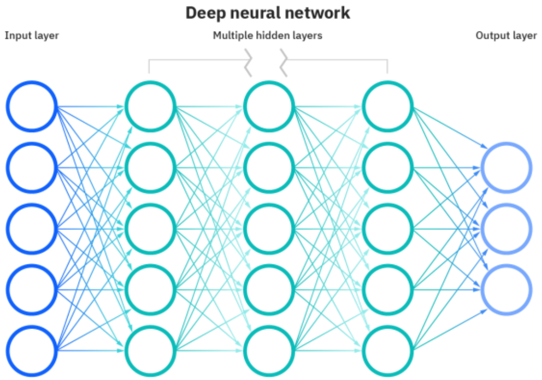
DEEP REINFORCEMENT LEARNING

Reinforcement Learning



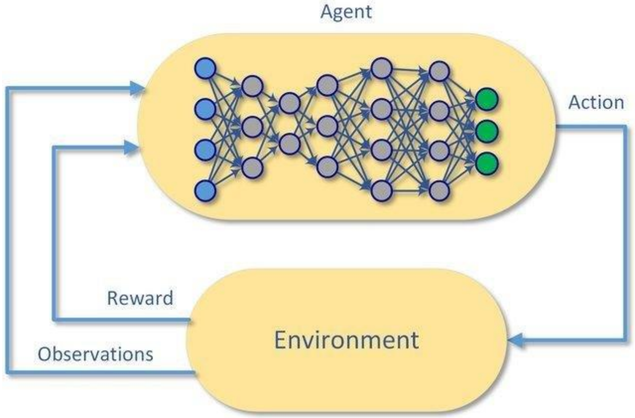
+

Deep Learning



=

Deep Reinforcement Learning



DEEP REINFORCEMENT LEARNING

Deep learning approximates function F by parametrizing function G using artificial neural networks

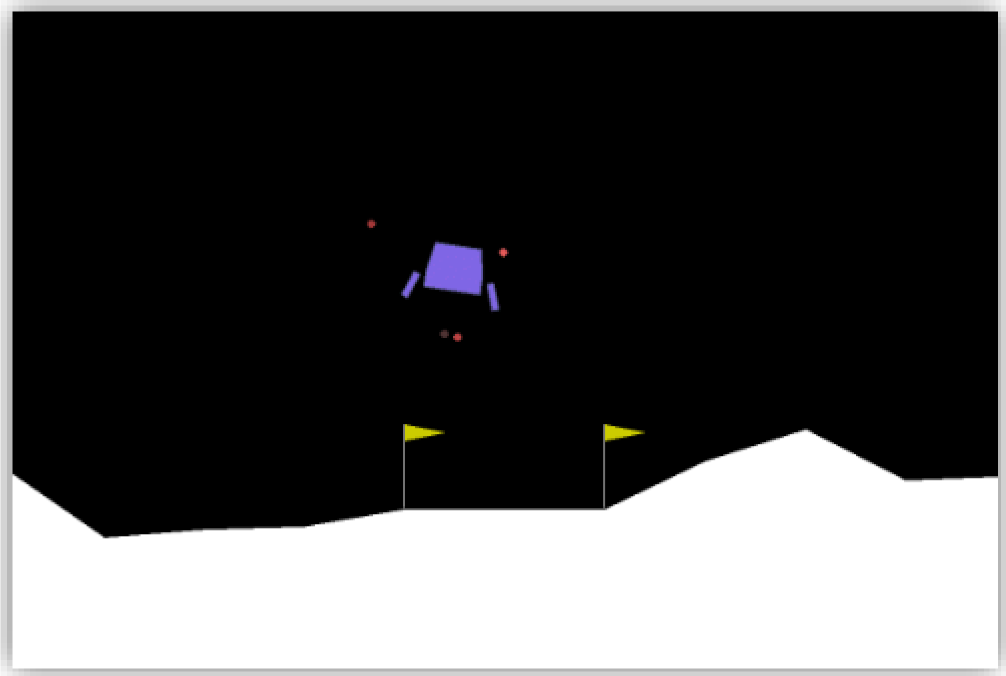
$$G(x;\theta) \text{ or } G_\theta(x) \approx F(x)$$

Where θ is neural network weights

Deep reinforcement learning uses neural networks to approximate its functions

$$Q_\theta(s_t, a_t) \approx Q^*(s_t, a_t)$$
$$V_\theta(s_t) \approx V^*(s_t)$$
$$\pi_\theta(s_t) \approx \pi^*(s_t)$$

OPENAI GYM



https://www.gymlibrary.dev/

LIBRARIES AND FRAMEWORKS FOR DEEP RL

Stable Baseline3: Reliable Reinforcement Learning Implementations

https://stable-baselines3.readthedocs.io/en/master/

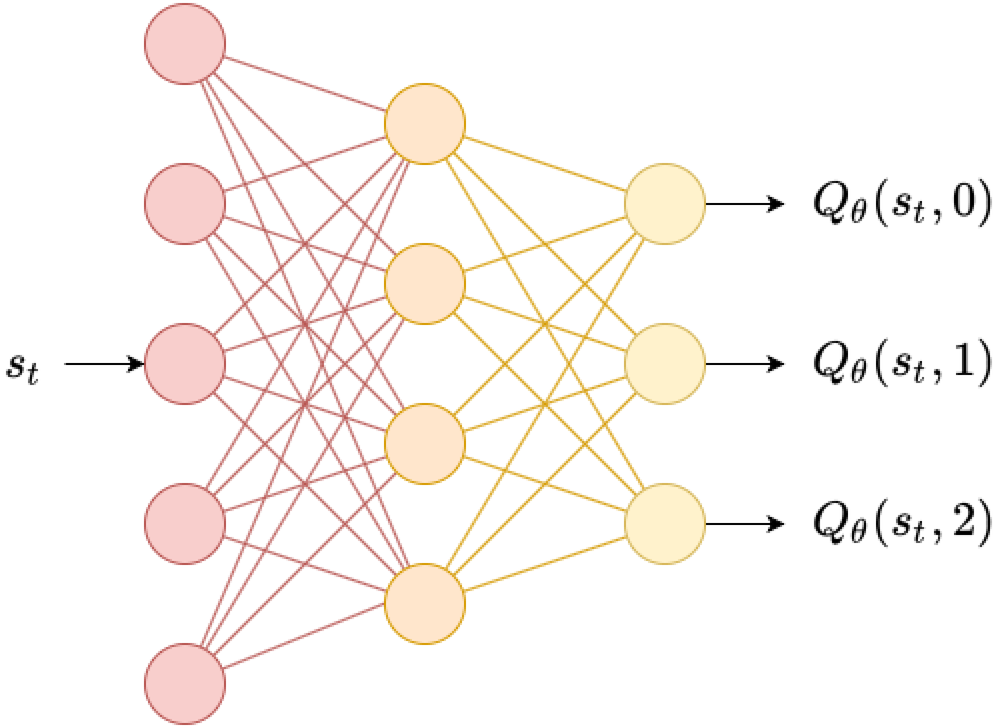
RLlib: Industry-Grade Reinforcement Learning

https://docs.ray.io/en/latest/rllib/index.html

LIBRARIES AND FRAMEWORKS FOR DEEP RL

STABLE BASELINE3 SAMPLE CODE

DQN CODE REVIEW



Q&A

THANK YOU

Alireza Nobakht

a.nobakht13@gmail.com

linkedin.com/in/alireza-nobakht/