EEA-EV002
Advanced topics in Reinforcement Learning

Overview

Syeda Sakira Hassan

Department of Electrical Engineering and Automation
Aalto University

January 12, 2022

Contents

Next

Overview of the course

Preview of RL

Deep RL

Teaching

Lectures

- 12.01-16.02
- Every Wednesday from 10.00-12.00.
- Zoom: `https://aalto.zoom.us/j/65356356800`

Sakira Hassan and Prof. Simo Särkkä

Presentations

- 5 sessions

Reports

- Summarizing the topics

Independent study

- Reading

Learning goals

- Understand the advanced Reinforcement Learning algorithms.
- Understand the application areas.
- Find a preference algorithm to solve Reinforcement Learning problems.

Prerequisites

- Basic knowledge of reinforcement learning
- Familiar with supervised learning methods
- Familiar with basic matrix algebra and optimization algorithms
- Familiar with deep neural networks
- Familiar with fundamentals of control-theory
- Familiar with Python and its libraries

Grading and evaluation

- The grading scale is pass/fail.
- One presentation based on the paper listed or can choose a related topic.
- Participation in every seminar is compulsory.
- One page (250-300 words) summary of each session.
- Become an opponent in at least one session.
- Grading assessment depends on technical correctness, writing quality, and language.

Workload

- Lectures 12 h
- Independent study 15 h
- Preparation to presentation work 30–40 h
- Reports 15 h
- Total 72–82 h

Materials

- Link to the papers (MyCourses)
- Sutton, R.S. and Barto, A.G., 2018. Reinforcement learning: An introduction. MIT press.
- Graesser, L. and Keng, W.L., 2019. Foundations of deep reinforcement learning: theory and practice in Python. Addison-Wesley Professional.
- Any relevant material.

Contents

- We focus on reinforcement learning algorithms based on deep neural network.
- List of topics: `https://mycourses.aalto.fi/course/view.php?id=34509`

Questions?

Next

Overview of the course

Preview of RL

Deep RL

Intro

Machine learning

- Supervised learning
- Unsupervised learning
- **Reinforcement learning**

Finding suitable actions to take in a given situation in order to maximize a reward.[1]

[1] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press

Overview

- Concerned with solving sequential decision-making problems.
- Real world problems: playing video games, sports, driving, optimizing inventory, robotic control
- Objective or goal, such as winning, arriving safely, minimizing inventory cost
- A subfield of AI that dates back to optimal control theory and Markov decision processes (MDPs).

Terminology

- Environment (S, A)
- Agent
- State $s_t \in S$
- Action $a_t \in A$
- Reward function $r_t = R(s_t, a_t, s')$
- Dynamics $p(s' \mid s_t, a_t)$
- Time horizon T
- Discount factor γ
- Policy $\pi(s_t, \theta) = a_t$
- Trajectory $\tau = (s_1, a_1, r_1), \ldots, (s_T, a_T, r_T)$

Let's play a game!

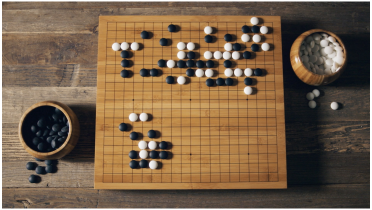


Figure: Alpha Go

- Environment?
- Agent?
- Reward?
- Action?

RL as MDP

What is MDP! An MDP is defined by a 5-tuple (S, A, P, R, γ)

$$s' \sim p(s' \mid s_t, a_t) \tag{1}$$

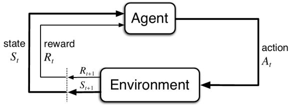


Figure: The agent-environment interaction in MDP [2].

Success stories of RL

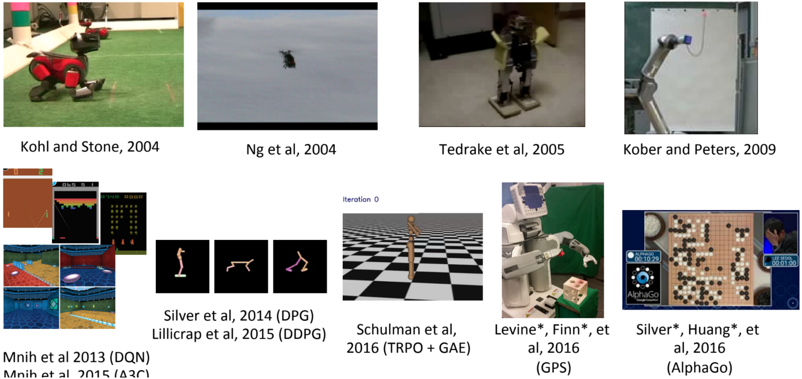


Figure: Image taken from Deep RL bootcamp.

RL inference: What an agent learns?

- A policy π: $a \sim \pi(s)$

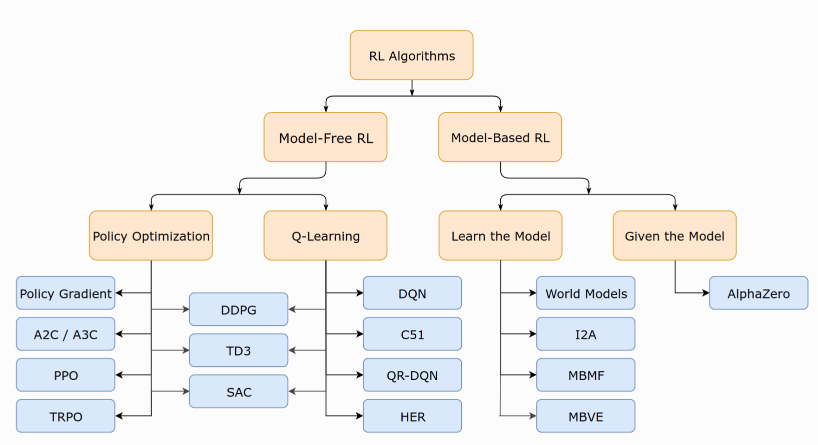
- A value function $V(s)$ or $Q(s, a)$

$$V_\pi(s) = \mathbb{E}_{s_0=s,\tau\sim\pi}\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_t = s\right] \tag{2}$$

$$Q_\pi(s,a) = \mathbb{E}_{s_0=s,a_0=a,\tau\sim\pi}\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_t = s, a_t = a\right] \tag{3}$$

- The environment model $P(s' \mid s, a)$

Algorithms [1]



Next

Overview of the course

Preview of RL

Deep RL

Challenges in Exact methods

Issues

- Curse of dimensionality
- Model

Approximation methods

- Estimate the value function
- Policy optimization
- **Function approximation**
- Model predictive control

Approximation methods I

Estimate the value function

- Monte carlo methods
- Bootstrapping (Temporal-difference learning)

Policy optimization

- Optimize the policy $\pi(a \mid s; \theta)$ directly
- Update $\theta \leftarrow \theta + \alpha \nabla_\theta J(\pi_\theta)$

Approximation methods II

Function approximation

- Estimate the value functions

$$V(s, w) \approx V_{\pi}(s) \tag{4}$$

$$Q(s, a, w) \approx Q_{\pi}(s, a) \tag{5}$$

- Many function approximators.
  - Linear combination of features
  - Neural networks
  - Decision trees
  - Nearest neighbors
  - Fourier/wavlet bases

Deep learning

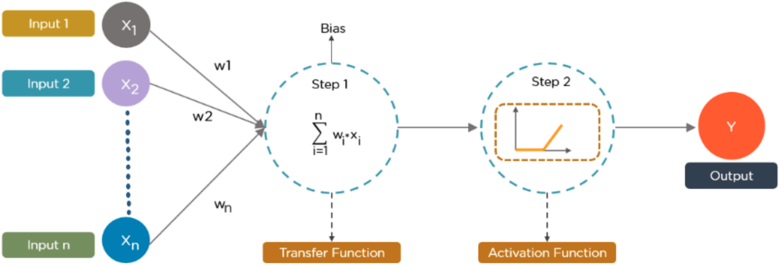


Figure: A simple neural network.

Environment Platforms

- OpenAI Gym
- Tensor Trade (Trading)
- VIZDoom (Game)
- Deepmind OpenSpiel (Game)
- Ns3 Gym (Networking)
- OpenSim (Biomechanics)
- AWS DeepRacer (Autonomous Vehicles)
- Many more ...

Libraries

- OpenAI Baseline
- RLLib
- CleanRL
- SLM Lab
- Stable baseline
- RLCodebase
- JaxRL
- DeepRL.jl
- Many more ...

References

GRAESSER, L., AND KENG, W. L.
Foundations of deep reinforcement learning: theory and practice in Python.
Addison-Wesley Professional, 2019.

SUTTON, R. S., AND BARTO, A. G.
Reinforcement learning: An introduction.
MIT press, 2018.

Thank you for listening!

Questions?