An Introduction to Reinforcement Learning

LINKÖPINGS UNIVERSITET

Farnaz Adib Yaghmaie

Linkoping University, *Sweden*
farnaz.adib.yaghmaie@liu.se

March 12, 2021

A few words about the RL workshop

- Given by Farnaz Adib Yaghmaie and Lennart Ljung.
    - Fredrik Ljungberg will assist me today.

- The workshop is a part of LINK-SIC.

- We define the basic concepts in RL: a bit of theory is involved.

- The presentation is open to discussion:
    - Write your question or
    - Unmute your microphone and ask!

- Run the code to understand better.

References

- **A Crash course on RL**, Farnaz Adib Yaghmaie, Lennart Ljung
    - Handout
    - Code repository
        https://github.com/FarnazAdib/Crash_course_on_RL

- **Reinforcement Learning: An introduction**, Sutton and Barto, 2018. Available free online

- **Videos by David Silver** https://youtu.be/2pWv7GOvuf0

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] Richard S Sutton & Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

How RL is different from other branches of ML?

- No supervisor; only a reward
- The action will effect subsequent data
- Dynamic data *vs.* Static data

Examples of RL

- Maneuvering helicopter
- Defeating Backgammon champion
- Surpassing human performance in playing many Atari games
- Making a humanoid robot walk
- Research and find many others!
- Find a successful application in your field!

Beating Atari champions [2]
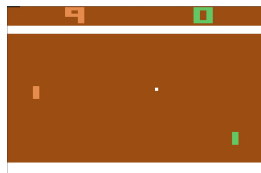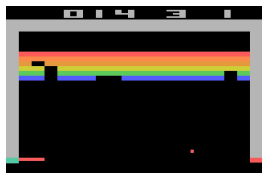
[2] Mnih et al. *Playing Atari with Deep Reinforcement Learning*, arXiv preprint, 2013.
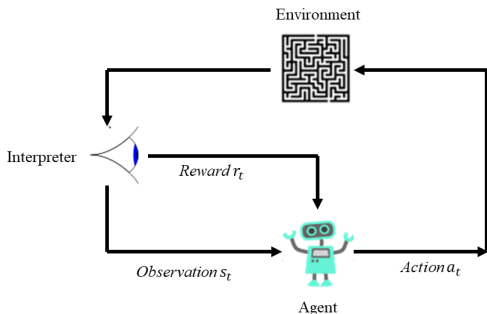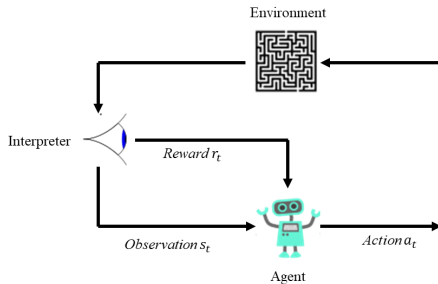
A graphical representation

Environment

Interpreter

Reward r_t

Observation s_t

Action a_t

Agent

Photo Credit: © https://en.wikipedia.org/wiki/Reinforcement_learning

Three main components in RL

- Reward
- Environment
- Agent

Environment

Interpreter

$Reward\ r_t$

$Observation\ s_t$

$Action\ a_t$

Agent

Reward r_t

- Reward is a scalar signal
- It indicates how well agent is doing at time t
- The aim is to maximize the total reward
- Or equivalently minimize a total penalty

Environment

- Generates dynamic data
  - Receives the action a_t
  - Emits a new state s_{t+1}
- Has unknown rules; i.e. dynamics
- Emits a reward r_t

Agent

- Receives the reward r_t
- Receives the observation s_t
- Execute an action a_t

  Generate actions to maximize the future rewards

Three main components of an RL agent

- Policy: The agent's rule to select action in a given state
- Value function: How good the agent's rule is
- Model of environment: The agent's interpretation of the environment

Not all are necessary!

Policy Gradient   *Learning policy*

Dynamic Programming based   *Learning value function*

Model building   *Learning the model of environment*

Let's play a game

You are playing chess with a
Grandmaster of chess, Dorsa.
What are

- Environment?
- Reward?
- Agent?

Photo Credit: @ https://www.pinterest.se/

Email your questions to

farnaz.adib.yaghmaie@liu.se