Reinforcement Learning

An Introductory Note

Jingye Wang

wangjy5@shanghaitech.edu.cn

Spring 2020

Contents

1 Introduction

Course Prerequisite:

- Linear Algebra
- Probability
- Machine Learning relevant course (data mining, pattern recognition, *etc*)
- PyTorch, Python

What is Reinforcement Learning and why we care:

a computational approach to learning whereby *an agent* tries to *maximize* the total amount of *reward* it receives while interacting with a complex and uncertain *environment*.[1]

Difference between Reinforcement Learning and Supervised Learning:

- Sequential data as input (*not i.i.d*);
- The learner is not told which actions to take, but instead must discover which actions yield the most reward by trying them;
- *Trial-and-error* exploration (balance between exploration and exploitation);
- There is *no supervisor*, only a reward signal, which is also *delayed*

Big deal: Able to Achieve Superhuman Performance

- Upper bound for Supervised Learning is human-performance.
- Upper bound for Reinforcement Learning ?

Why Reinforcement Learning works now?

- Computation power: many GPUs to do trial-and-error rollout;
- Acquire the high degree of proficiency in domains governed by simple, known rules;
- End-to-end training, features and policy are jointly optimized toward the end goal.

Sequential Decision Making:

- Agent and Environment: the agent learns to interact with the environment;
- Rewards: a scalar feedback signal that indicates how well agent is doing;
- Policy: a map function from state/observation to action models the agent’s behavior;
- Value function: expected discounted sum of future rewards under a particular policy;
- Objective of the agent: selects a series of actions to maximize total future rewards;
- History: a sequence of observations, actions, rewards;
- Full observability: agent directly observes the environment state, formally as Markov decision process (MDP);

- Partial observability: agent indirectly observes the environment, formally as partially observable Markov decision process (POMDP)

All goals of the agent can be described by the maximization of expected cumulative reward.

Types of Reinforcement Learning Agents based on What the Agent Learns

- Value-based agent:
  - Explicit: Value function;
  - Implicit: Policy (can derive a policy from value function);
- Policy-based agent:
  - Explicit: policy;
  - No value function;
- Actor-Critic agent:
  - Explicit: policy and value function.

Types of Reinforcement Learning Agents on if there is model

- Model-based:
  - Explicit: model;
  - May or may not have policy and/or value function;
- Model-free:
  - Explicit: value function and/or policy function;
  - No model.