Introduction

Reinforcement Learning

2019

Sepehr Maleki

University of Lincoln
School of Engineering

The Reinforcement Learning Problem

- We learn by interacting with our environment.
- Nature has no explicit teacher but we have a direct sensorimotor connection to our environment.
- Exercising this connection produces a wealth of information about cause and effect, about the consequences of actions, and about what to do in order to achieve goals.
- Learning from interaction is a foundational idea underlying nearly all theories of learning and intelligence.
- In this course we explore a computational approach to learning from interaction.

Reinforcement Learning

- Reinforcement learning problems involve learning how to map situations to actions so as to maximise a numerical reward signal.
- They are closed-loop problems because the learning system's actions influence its later inputs.
- The learner is not told which actions to take, as in many forms of machine learning, but instead must discover which actions yield the most reward by trying them out.
- Actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards.

(S. Maleki 2019)

Interaction With Environment

The learner and decision-maker is called the *agent*. The thing it interacts with, comprising everything outside the agent, is called the *environment*.

- An agent must be able to sense the **state of the environment** to some extent (through sensors like camera, etc.).
- An agent must be able to take **actions** that affect the state.
- The agent also must have a **goal or goals** relating to the state of the environment.

(S. Maleki 2019)

Exploration vs Exploitation

- One of the challenges that arise in reinforcement learning is the trade-off between exploration and exploitation.
- To obtain a lot of reward, an agent must prefer actions that it has tried in the past and found to be effective in producing reward.
- But to discover such actions, it has to try actions that it has not selected before.
- The agent has to exploit what it already knows in order to obtain reward, but it also has to explore in order to make better action selections in the future.
- The dilemma is that neither exploration nor exploitation can be pursued exclusively without failing at the task.It must try a variety of actions and progressively favour those that appear to be best.

(S. Maleki 2019)

Examples

Some examples and possible applications that have guided development of RL:

- A master chess player makes a move.

- A gazelle calf struggles to its feet minutes after being born. Half an hour later it is running at 20 miles per hour.

- A mobile robot decides whether it should enter a new room in search of more trash to collect or start trying to find its way back to its battery recharging station. It makes its decision based on the current charge level of its battery and how quickly and easily it has been able to find the recharger in the past.

Elements of Reinforcement Learning

- **Policy**: A policy defines the learning agent's way of behaving at a given time.
- **Reward**: A reward signal defines the goal in a reinforcement learning problem. On each time step, the environment sends the agent a single number, a reward. The agent's sole objective is to maximise the total reward it receives over the long run.
- **Value Function**: A value function specifies what is good in the long run. Roughly speaking, the value of a state is the total amount of reward an agent can expect to accumulate over the future, starting from that state.
- **Model**: A model the behaviour of the environment, or more generally, allows inferences to be made about how the environment will behave.

(S. Maleki 2019)