Reinforcement Learning

SLIDES BASED ON A BOOK OF THE SAME NAME BY SUTTON AND BARTO.

Definition

Reinforcement learning is goal-directed learning and decision-making.

It is distinguished from other approaches by:

- emphasis on learning from direct interaction with its environment,
- not relying on exemplary supervision or complete models of the environment

Motivation

People learn by interacting with their environment.

Examples:

- An infant that plays, waves its arms, or looks about, has no explicit teacher, but it does have a direct sensorimotor connection to its environment.
- When learning to drive a car, we get feedback from the car and sometimes from fellow drivers.
- When holding a conversation, we are aware of how our environment responds to us.

Characterization

Reinforcement learning is learning what to do so as to maximize a numerical reward signal.

The learner is not told which actions to take.

They instead must discover which actions yield the most reward by trying them.

Most of the time, actions may affect not only the immediate reward but also subsequent rewards.

Exploration vs. Exploitation

One of the challenges that arise in reinforcement learning is the trade-off between:

- exploration and
- exploitation.

To maximize reward, a reinforcement learning agent should prefer actions that it has tried in the past and found to be rewarding.

However, to discover such actions, it has to try actions that it has not selected before.

Exploration vs. Exploitation

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.

Problem vs. Method

Reinforcement learning explicitly considers the *whole* problem.

This is in contrast with many approaches that consider sub-problems and then put them together.

Example

In chess, there may be immediate rewards to a move.

Some of the rewards may even be negative.

However, a chess player considers the overall strength of their situation.

Elements

Beyond the agent and the environment, there are four main sub-elements of a reinforcement learning system, a:

- *reward function*,
- *value function*,
- *policy* and, optionally,
- *model* of the environment.

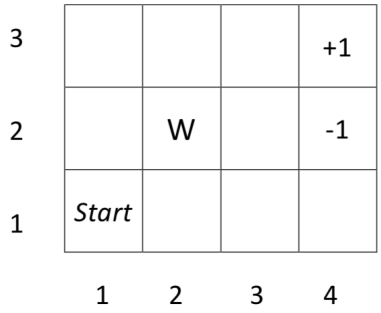
Example World

Agent lives in a grid

Small reward for each step

Big rewards come at end

Goal: Maximize sum of rewards



Reward Function

Small reward for each step: -0.04

Big rewards come at end: +1 and -1.

3	-0.04	-0.04	-0.04	+1
2	-0.04	W	-0.04	-1
1	*Start*	-0.04	-0.04	-0.04
	1	2	3	4

Value Function

A reward function indicates what is good in an immediate sense.

A *value* or *utility function* specifies what is good in the long run.

The *value* of a state represents the total amount of reward an agent can expect to accumulate from that state.

Value Function

A state may yield a low immediate reward, as seen in the earlier table.

However, since such a state may lead to one with a high reward, this should be captured in the value of the such a state.

Value Function

Below is a value function, sometime also called a utility function.

3	0.812	0.868	0.918	+1
2	0.762	W	0.660	-1
1	0.705	0.655	0.611	0.388
	1	2	3	4

Policy

A *policy* defines the learning agent's way of behaving at a given time.

A policy is a mapping from states to actions.

Policy

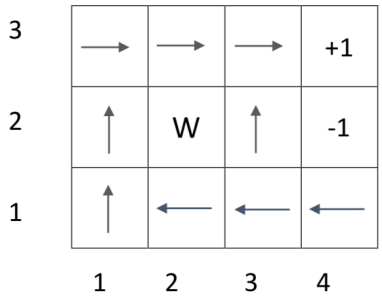
The policy is the core of a reinforcement learning agent.

It is used to determine the behavior of an agent.

A policy depends on the rewards which in turn determine the value function.

Policy

Below is a policy that was derived from the value function seen earlier (and replicated on the right.)



3	0.812	0.868	0.918	+1
2	0.762	W	0.660	-1
1	0.705	0.655	0.611	0.388
	1	2	3	4