celtra

Reinforcement Learning

Tom Vodopivec

IADS Big Data and Analytics Summer School
2019-08-05

The Field

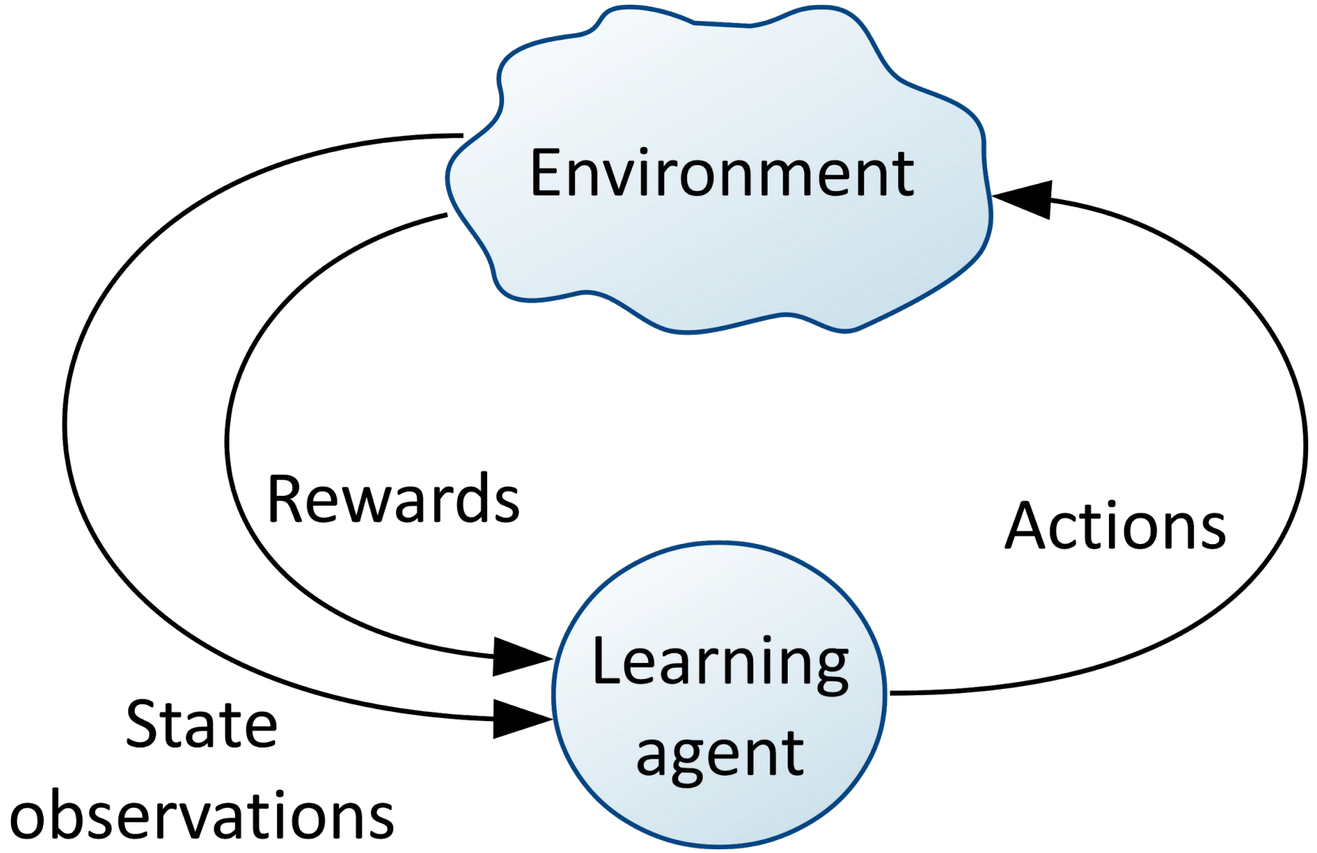
Artificial Intelligence > Machine Learning

(A special case of) supervised learning

+

But also has elements of unsupervised learning

The Reinforcement Learning Setting



The Reinforcement Learning Setting

Agent

Behaviour > Actions

+

Environment

States

Rewards > define the goal of the agent

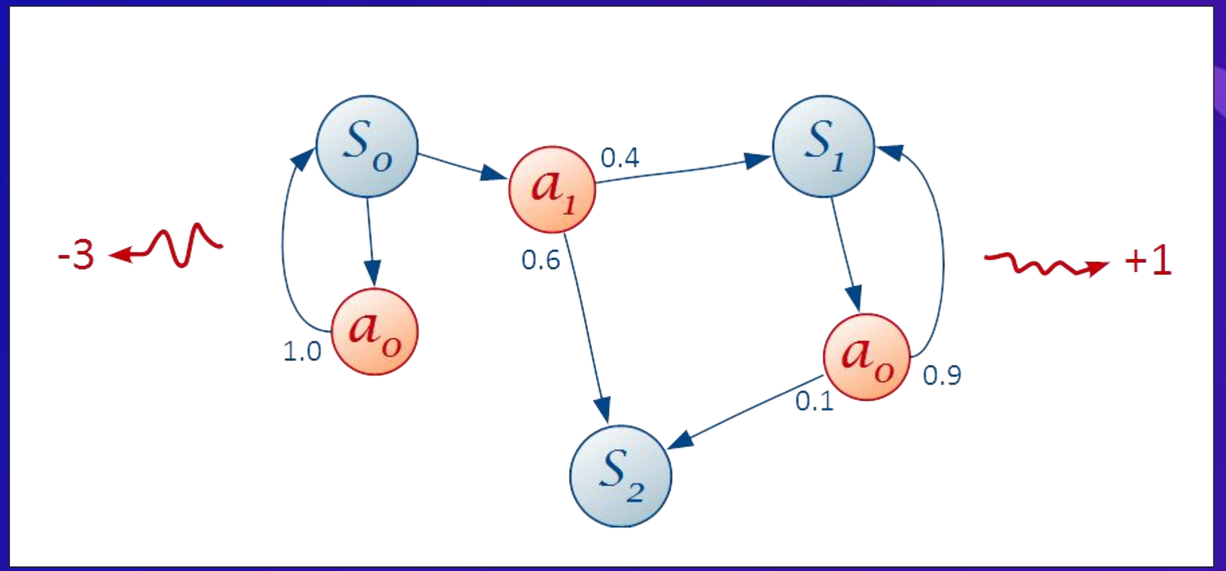
Markov Decision Processes

States

Actions

Rewards

Transition probabilities



Markov Decision Processes

Trajectory (concept of time)

(Expected) return

Discount rate

+

Continuing vs. episodic tasks

Terminal states

Non-stationary MDPs

Partially-observable MDPs

What to Learn From?

Experience = Samples of interaction with environment

- Real (learning)
- Simulated (planning)

+

Exploration-exploitation trade-off

Uncertainty

How to Store Knowledge?

State-value function

Action-value function

Behaviour policy

Model (of the environment)

+

Representations

- Tabular vs. approximate
- Fixed vs. adaptive

Updating Knowledge = Learning

Goal: learn optimal policy (and optimal value function)

Generalized policy iteration

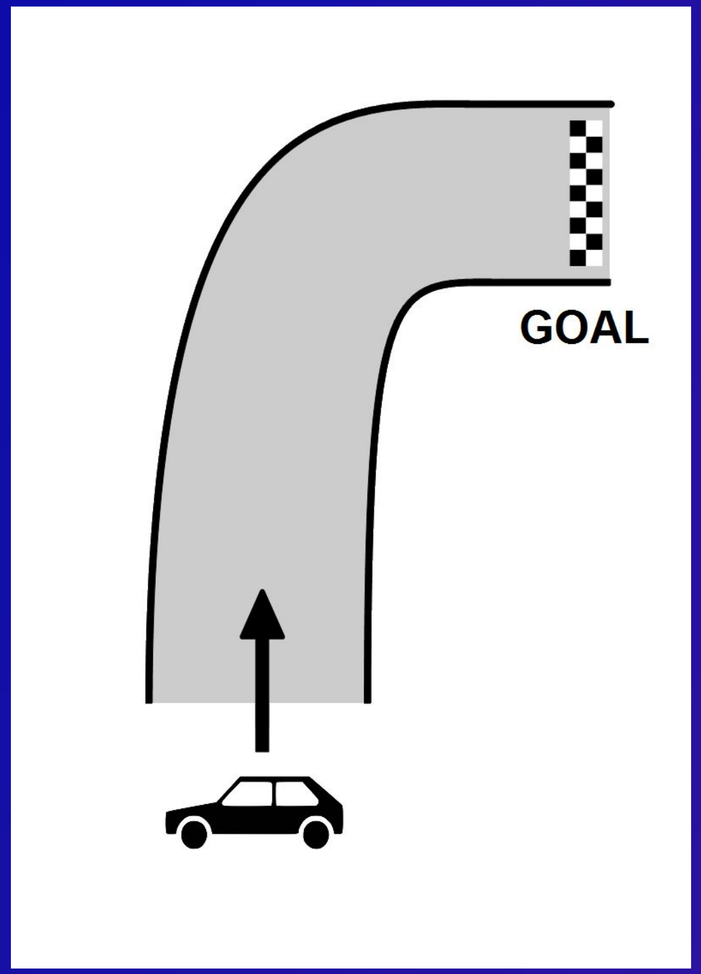
- Policy evaluation
- Policy improvement

+

Algorithms

- Model-based
- Model-free

Race Track Example



Algorithms

Dynamic programming

Monte Carlo

+

Temporal-difference learning

- Bootstrapping
- Eligibility traces

Algorithms

Dynamic programming: policy iteration, value iteration

Monte Carlo

+

Temporal-difference learning: TD(λ), Sarsa, Q-learning, Deep Q-learning

- Bootstrapping
- Eligibility traces

Algorithms

Dynamic programming: policy iteration, value iteration

Monte Carlo

+

Temporal-difference learning: TD(λ), Sarsa, Q-learning, Deep Q-learning

- Bootstrapping
- Eligibility traces

Policy gradient and Actor-Critic: REINFORCE

Applications

Games: AlphaGo

Scheduling tasks: optimization of memory control

Modelling bird movement

Web services / optimization

+

Reading Material

Quick and practical state-of-the-art:

Thomas Simonini, Deep Reinforcement Learning course with Tensorflow

Most comprehensive and best foundations:

Richard S. Sutton and Andrew G. Barto, Reinforcement Learning, An introduction, second edition

Outstanding applications:

Silver et al., AlphaGo, AlphaGo Zero, AlphaZero

Recap

Agent, actions, environnement, state, rewards

Representation of knowledge

Explore and collect experience

Update knowledge = learn

Improve behaviour

+