

AIC – Reinforcement Learning

Freek Stulp and Michèle Sebag

Université Paris-Saclay

Reinforcement Learning: What and Why?

What? Learn to perform the best actions, given only rewards and penalties

- Useful for training dogs



Reinforcement Learning: What and Why?

What? Learn to perform the best actions, given only rewards and penalties

- Useful for training dogs
- Enable agents/robots to learn tasks autonomously



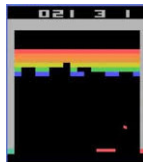
TD-Gammon

Tesauro et al.



Flip pancake

Kormushev et al.



Atari games

Google Deepmind

Reinforcement Learning: What and Why?

What? Learn to perform the best actions, given only rewards and penalties

- Useful for training dogs
- Enable agents/robots to learn tasks autonomously

Why? Advantage for human

- Does not need to specify the optimal behavior (difficult)
- Must only the reward function (much easier)



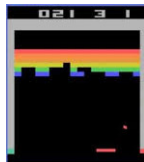
TD-Gammon

Tesauro et al.



Flip pancake

Kormushev et al.



Atari games

Google Deepmind

Reinforcement Learning: What and Why?

You (the agent or robot) learn from interaction with the world.

The observations that you make depend on the actions that you take.

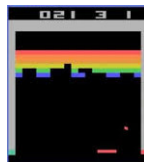
So in reinforcement learning, you make your *own* big data!



TD-Gammon
Tesauro et al.



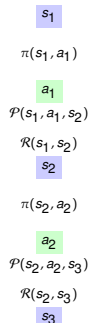
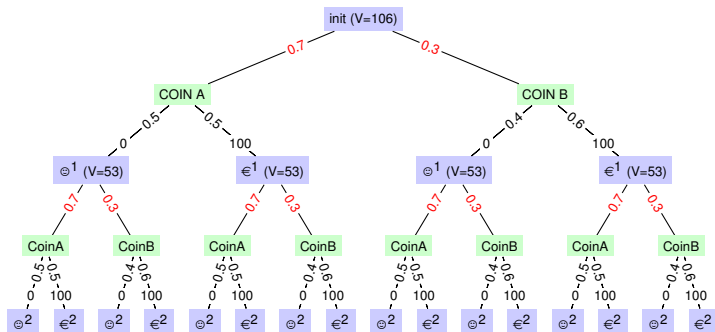
Flip pancake
Kormushev et al.



Atari games
Google Deepmind

Topics of the Course

- Markov Decision Process
 - Underlying model of RL problems
- Dynamic Programming
 - Recursive algorithm for environments with known dynamics
- Algorithms for solving discrete RL problems
 - Monte Carlo methods and Temporal Difference learning
- Algorithms for solving continuous RL problems
 - Function approximation and direct policy search
- And beyond: deep learning for RL, inverse RL, RL for robotics



Style of the Course: Constructivist Didactics

- Lectures (theory) and programming (practice) both important
- Implementing algorithms \equiv understanding algorithms
 - agent influences observations: study the dynamics of learning
- Subsequent lectures build on knowledge acquired through implementation

Style of the Course: Constructivist Didactics

- Lectures (theory) and programming (practice) both important
- Implementing algorithms \equiv understanding algorithms
 - agent influences observations: study the dynamics of learning
- Subsequent lectures build on knowledge acquired through implementation

*"I hear and I forget.
I see and I remember.
I do and I understand."
Confucius*

Style of the Course: Constructivist Didactics

- Lectures (theory) and programming (practice) both important
- Implementing algorithms \equiv understanding algorithms
 - agent influences observations: study the dynamics of learning
- Subsequent lectures build on knowledge acquired through implementation

*"I hear and I forget.
I see and I remember.
I code Python and I understand."*
Confucius

Reinforcement Learning and Life

Life is also very much about choosing the right actions



Reinforcement Learning and Life

Life is also very much about choosing the right actions

⇒ Many parameters in RL correspond to real-life human strategies



Reinforcement Learning and Life

Life is also very much about choosing the right actions

⇒ Many parameters in RL correspond to real-life human strategies

Discount factor γ

$\gamma \approx 0$ hedonists

$\gamma \approx 1$ people that worry about the future a lot

Reinforcement Learning and Life

Life is also very much about choosing the right actions

⇒ Many parameters in RL correspond to real-life human strategies

Discount factor γ

$\gamma \approx 0$ hedonists

$\gamma \approx 1$ people that worry about the future a lot

Exploration parameter ϵ

$\epsilon = 1$ people that always try new things

$\epsilon = 0$ people that always go to the same restaurant

Reinforcement Learning and Life

Life is also very much about choosing the right actions

⇒ Many parameters in RL correspond to real-life human strategies

Discount factor γ

$\gamma \approx 0$ hedonists

$\gamma \approx 1$ people that worry about the future a lot

Exploration parameter ϵ

$\epsilon = 1$ people that always try new things

$\epsilon = 0$ people that always go to the same restaurant

Exploration decay parameter β

$\beta = 0.7$ people that are already boring at 25

$\beta = 1.0$ elderly people that are still cool!

Thank you for your attention! Any questions?

Figure : Ball-in-cup skill, which the Meka acquired through reinforcement learning

