

Intro to MDP and RL

Joseph Le Roux

22/06/26



1

Introduction

1 AI for Autonomous Agents?

How to take a sequence of decisions to solve a task!

1. What is the best sequence of decisions?

- This is an instance of *optimal control* based on optimisation (linear programming...)
- With the addition of random events (model failures, weather conditions, opponents...)

2. Interesting framework for Artificial Intelligence (AI)

- How can the agent learn to make good decisions from experience

3. Important Applications

- transports/factory labor/agriculture/health/chatbots...
- with possibly strong social impact

1. Hardware is more and more available, and it gets cheaper
 - specialised hardware
 - generic hardware (raspberry, jetson, arduino...)
 - cheap/portable GPUs (for neural networks)
2. Software is the limiting factor
 - cf. videos of tele-operated robots
3. For this kind of software, we need:
 - unconstrained/constrained optimisation (cf. Maths for engineering and this year OL/OC)
 - probability/analysis/numerical methods (cf. last year INFO1)
 - machine learning/neural networks (cf. this year and next year INFO2/INFO3)

1 Hardware/Software: Progress in the Last 15 Years (1)

- **tele-operated robots** Hardware gets better and better, and cheaper

PR1 (2007, \$400000) <https://youtu.be/dFFXgZENpDI>

Blue (2019, \$5000) <https://youtu.be/KZ88hPgrZzs>

Shadow (2021) <https://youtu.be/J-3CcyF4Emw>

- **autonomous vehicles** (big field!) Tremendous improvements, but real city traffic is harder than racing in the desert!

DARPA challenge 2005 <https://youtu.be/7a6GrKq0xeU>

Stanford Copter, 2008 <https://youtu.be/M-QUkgk3HyE>

Stanford car, 2019 <https://youtu.be/3x3SqeSdrAE>

Google car, 2011 <https://youtu.be/cdgQpa1pUUE>

1 Hardware/Software: Progress in the Last 15 Years (2)

- [robot worker](#) Compare indoor parkour and outdoor walk on snowy ground

Demo Parkour <https://youtu.be/tF4DML7FIWk>

Into the Wild <https://youtu.be/rV1hMGQgDkY>

- [agents for simulations/games](#) Purely software, these are famous achievement of RL for complex decision making

atari games <https://youtu.be/V1eYniJ0Rnk>

backgammon <https://youtu.be/RvgEA1XJiY8>

Go <https://youtu.be/MgowR4pq3e8>

- [Chatbots/Assistants](#) RL is used to fine-tune LLMs in order to distinguish useful from useless responses

RL for LLMs <https://youtu.be/Bqe5ERIZ9D4>



2

Roadmap

MDP (sess. 1/2)

- Agent-environment model
- Values , gains, policies
- Bellman equation: link local decision and future gains
- Tabular algorithms for optimal policy

Function Approximation (sess. 4/5)

Parametric models: MDP meets RL

- MC and TD with functions
- The case of linear functions

Monte-Carlo Estimation and Temporal Difference Methods (sess. 3)

MDP not known in advance (real-word)

- MC: approximate expectations
- TD: mix MC and Bellman

Policy Gradient (sess. 5/6)

Learn stochastic policy from data:

- Links with Likelihood Estimation
- Derivation of the gradient descent
- Variants (Actor-Critic)