ID de Contribution: **1** Type: **Non spécifié**

Talk by Yann Ollivier

jeudi 16 décembre 2021 14:00 (1h 30m)

This is the conference talk. Yann Ollivier will also give a colloquim talk the next day.

Abstract: Markov decision processes are a model for several artificial intelligence problems, such as games (chess, Go…) or robotics. At each timestep, an agent has to choose an action, then receives a reward, and then the agent's environment changes (deterministically or stochastically) in response to the agent's action. The agent's goal is to adjust its actions to maximize its total reward. In principle, the optimal behavior can be obtained by dynamic programming or optimal control techniques, although practice is another story.

Here we consider a more complex problem: learn all optimal behaviors for all possible reward functions in a given environment. Ideally, such a "controllable agent" could be given a description of a task (reward function, such as "you get +10 for reaching here but -1 for going through there") and immediately perform the optimal behavior for that task. This requires a good understanding of the mapping from a reward function to the associated optimal behavior.

We will present our recent theoretical and empirical results in this direction. There exists a particular "map" of a Markov decision process, on which near-optimal behaviors for all reward functions can be read directly by an algebraic formula. Moreover, this "map" is learnable by standard deep learning techniques from random interactions with the
environment.