Making Deep Q-learning Approaches Robust to Time Discretization

Corentin Tallec   Léonard Blier   Yann Ollivier

Université Paris-Sud, Facebook AI Research

June 4, 2019

What happens when **using standard RL methods** with **small time discretization** or **high framerate**?

- Usual RL algorithm + high framerate → failure

- Scalability limited by **algorithms**!
  Better hardware, sensors, actuators → Worse performance

- Contributes to **lack of robustness** of Deep RL:
  New environment → different framerate → new hyperparameters.

Low FPS   High FPS

Why is near continuous Q-learning failing?

There is no continuous time Q-learning

- As $\delta t \to 0$, $Q^\pi(s, a) \to V^\pi(s)$

- Q^π **does not depend on actions when** $\delta t \to 0$
  $\implies$ Cannot use Q^π to select actions!

There is no continuous time ε-greedy exploration

- ε-greedy, $\varepsilon = 1$ pendulum:

$$\delta t = .05 \qquad\qquad \delta t = .0001$$

YES

To know how:

Poster #32 this evening

Low FPS	High FPS