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.

<table>
<thead>
<tr>
<th>Low FPS</th>
<th>High FPS</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
</tr>
</tbody>
</table>
Why is near continuous Q-learning failing?

There is no continuous time Q-learning

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

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

There is no continuous time $\varepsilon$-greedy exploration

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

  $\begin{array}{c|c|c|c}
  \delta t &=& .05 & \delta t = .0001 \\
  \end{array}$
Can we solve this?

YES

To know how:

Poster #32 this evening

Low FPS  High FPS