Reinforcement Learning project

Hugo Cadet
Mathieu Martial
Victoire Gorge

# Visualization of eligibility traces: comparison with TD and MC

**November 2023**

## CONTEXT

In regular learning methods like Monte Carlo (MC), the agent needs a lot of examples to figure things out, and it has to wait until the end of a task to know if it did well. Temporal Difference (TD) tries to solve this by guessing future rewards instead of waiting for the final result. However, TD can have problems with fitting too closely to the data and dealing with long-term consequences.

Eligibility traces, a mix of both methods, offer a smarter and more flexible approach. They work like a memory for the agent, keeping track of past actions and their outcomes. This helps the agent learn better, especially in situations where the effects of its actions take time to show up. The "trace" **e(s,a)** is a score that each past state-action pair gets, guiding the agent's learning process. This makes eligibility traces a useful tool in environments with delayed consequences.

## ADVANTAGES

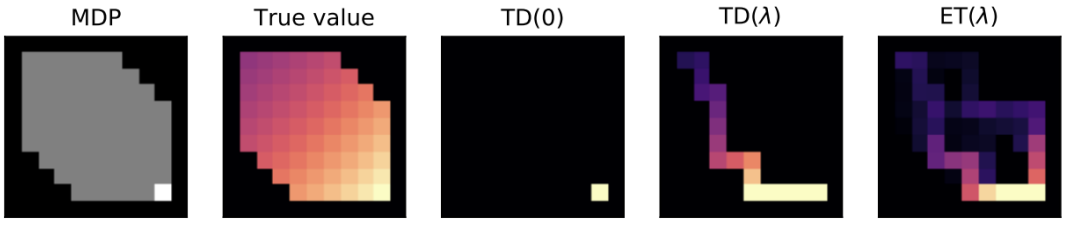
➔ they adeptly manage **non-stationary** and **delayed rewards**, because the agent can adjust its policy by considering anticipated future rewards and the record of past experiences.

➔ they prove effective in addressing challenges involving **continuous states** and actions, enabling the agent to refine its policy with the gradient of the value function.

➔ they are more effective than TD (empirically)

## Tuto : Representation of the trace for the comparison

Reproduce those graphs (from *arXiv:2007.01839v1 [cs.LG] 3 Jul 2020*) in an animated way.




Compare the accumulating/ replacing traces.