POLITECNICO
MILANO 1863

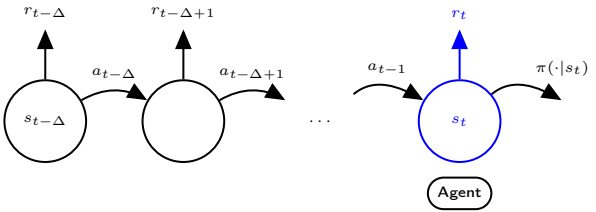
Delayed Reinforcement Learning by Imitation

Pierre Liotet Davide Maran Lorenzo Bisi Marcello Restelli

July 2022
Thirty-ninth International Conference on Machine Learning (ICML-22)

Problem Statement

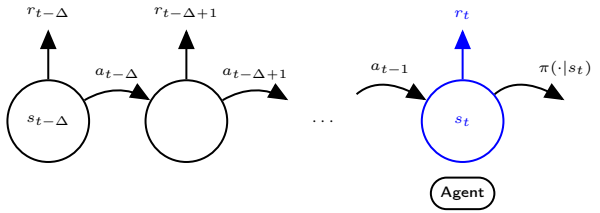
- Reinforcement learning:



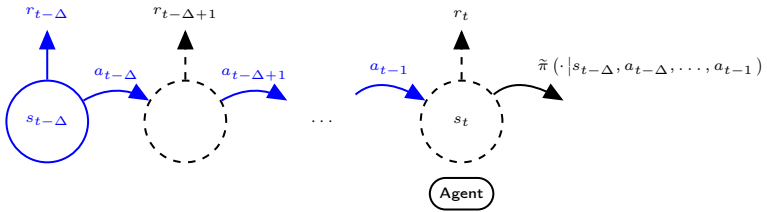
- Delay:

Problem Statement

- Reinforcement learning:



- Delay:

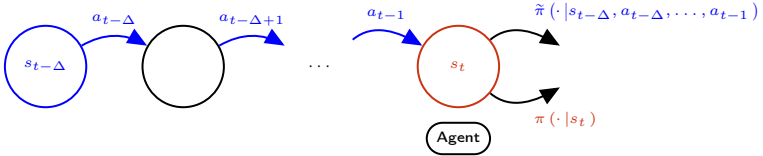


Challenges

- Complications:
  - Non-Markovianity
  - Augmented state space
  - Performance loss
  - Non integer delays

Simple yet Efficient Solution: Imitation Learning

- Duality of trajectories

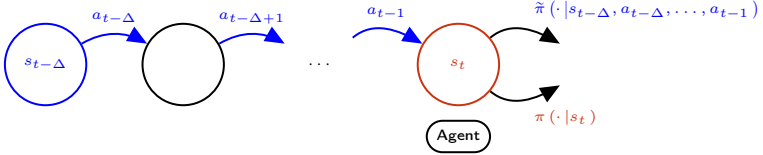


- Imitation learning with DAgger (Ross et al., 2011):

$$\widetilde{\pi}\left(a\,|s_{t-\Delta}, a_{t-\Delta,\dots,a_{t-1}}\right) = \int_{\mathcal{S}} b\left(s\,|s_{t-\Delta}, a_{t-\Delta}, \dots, a_{t-1}\right))\pi(a|s)ds.$$

Simple yet Efficient Solution: Imitation Learning

- Duality of trajectories



- Imitation learning with DAgger (Ross et al., 2011):

$$\widetilde{\pi}\left(a\,\middle|s_{t-\Delta}, a_{t-\Delta,\dots,a_{t-1}}\right) = \int_{\mathcal{S}} b\left(s\,|s_{t-\Delta}, a_{t-\Delta}, \dots, a_{t-1}\right))\pi(a|s)ds.$$

Theoretical Analysis

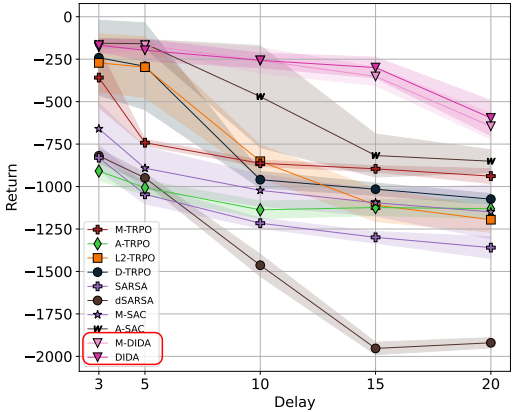
- Bounding performance loss:

$$\underbrace{\mathop{\mathbb{E}}_{s\sim b(\cdot|x)}[V^{\pi}(s)]}_{\substack{\text{undelayed expert}\\ \text{performance}}} - \underbrace{V^{\tilde{\pi}}(x)}_{\substack{\text{delayed}\\ \text{performance}}} \leqslant \frac{C}{1-\gamma} \underbrace{\mathop{\mathbb{E}}_{x'\sim d_x^{\tilde{\pi}}(\cdot)}\left[\sqrt{\mathop{\mathbb{V}\mathrm{ar}}_{s\sim b(\cdot|x')}(s|x')}\right]}_{\substack{\text{measure of uncertainty}\\ \text{on the current unobserved state}}}.$$

- Deterministic process $\implies$ bound $= 0$

Experiments

- Mean return as a function delay



Experiments

- Robotic locomotion - `mujoco` (Todorov et al., 2012)
- Trading

References

S. Ross, G. Gordon, and D. Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011.

E. Todorov, T. Erez, and Y. Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033. IEEE, 2012.