

Reinforcement Learning Project

Houda Alberts & Urja Khurana & Lina Murady & Caitlin Lagrand

November 2018

1 Topic

The exploration problem can be made a lot easier if a good solution is demonstrated [1]. Study the proposed solution: Can demonstrations be misleading? How good do demonstrations need to be? Does that depend on properties of the environment? Consider e.g. the ‘chain’ MDP in [2].

2 Experiments

We would like to apply the same algorithm to different environments and demonstrations. These environments and demonstrations are described below.

- Different environments
 - Finite state space (not much) & Discrete actions (not much): FrozenLake
 - Finite state space (not much) & Discrete actions (a lot): Roulette
 - Finite state space (not much) & Continuous actions: MountainCarContinuous
 - Finite state space (a lot) & Discrete actions (not much): FreewayRAM
 - Infinite state space & Discrete actions (not much): CartPole
 - Infinite state space & Continuous actions: Pendulum

Strictly speaking, MountainCarContinuous has an infinite state space, but we had a hard time to find an environment with a finite state space and a continuous action space. We were wondering if we could still use it, either by discretising it, or assuming that not many different states exists, since it is limited to only two coordinates (x and y) within the two mountain tops.

- Different demonstrations
 - Perfect demonstrations: show the perfect demonstration
 - Inefficient demonstrations: show a non optimal demonstration
 - Rough demonstrations: show a demonstration that gets close to the goal, but does not reach it

References

- [1] Tim Salimans and Richard Chen. Learning montezuma’s revenge from a single demonstration. *OpenAI Blog*, 2018.
- [2] Nikos Vlassis and Marc Toussaint. Model-free reinforcement learning as mixture learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1081–1088. ACM, 2009.