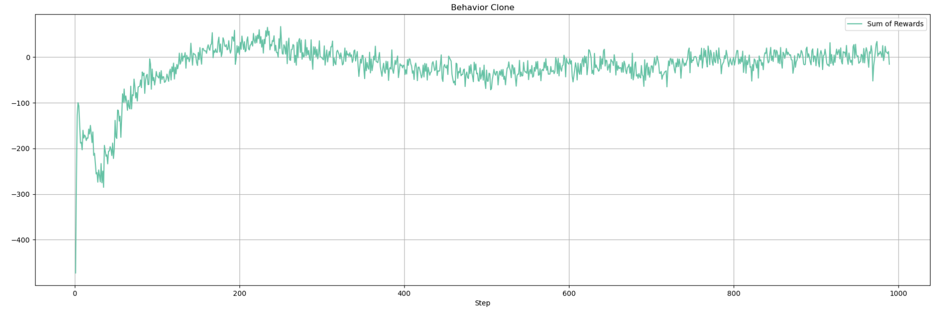
TME #11: Imitation Learning

In Imitation Learning, we try to learn a policy that is similar to the one of an expert agent, for which we have access to a set of realised transitions ϵ.

Behavior Cloning

Behavior Cloning tackled the problem by directly maximizing the log-likelihood of the expert transitions samples under the learned policy. We therefore successively take several gradient steps on batches of our expert data. The observed transitions in the environment are not stored, and not considered in the gradient update:

Figure Sum of Cumulative Rewards, Computed every 100 iterations on a mean of 100 episodes.

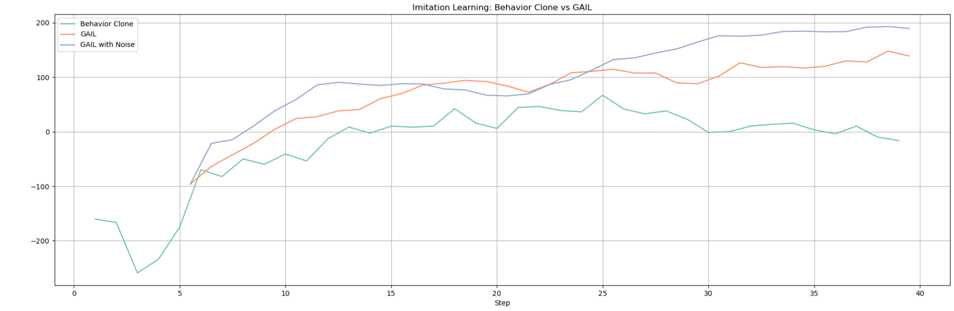


Capitalizing on the expert knowledge, the agent is able to learn a relatively efficient policy and reach a positive sum of rewards aroud step 200 (of test mode). But it then stays stuck, and is unable pass 0 again: learning strictly from the expert data does know allow our agent to generalize well to the states encountered in the environment.

GAIL

We implement the Generative Adversarial Imitation Learning (GAIL) algorithm with a clipped PPO update. To stabilize the optimization, rewards are clipped to the $[-100, 0]$ interval. We further run a second experiment with the same GAIL model with noisy update, where a random noise $\epsilon\ \mathcal{N}(0, 0.01)$ is added to transition pairs before each step update of the discriminator. As expected, the GAIL architetures by far outperform the Behavior Clone agent, with noisy version reaching 200 in cumulative rewards. Without noise, the probabilities output by the Discriminator when fed expert (resp. collected) transitions converges towards 1 (resp. 0) after around 250K steps of gradient. With noise on the other hand, they converge towards 0.85 (resp. 0.1). This indicates that adding noise had helped us in learning a policy that generates transitions similar enough to the ones of the expert that the Discriminator is fooled on 10% of them.

Figure Sum of Cumulative Rewards, Computed every 1000 iterations on a mean of 100 episodes. Smoothed on 10 points.


Imitation Learning: Behavior Clone vs GAIL
Behavior Clone
GAIL
GAIL with Noise
Step