

Week 8 - Generative Adversarial Imitation Learning (GAIL)

Submitted by hollygrimm on Mon, 07/30/2018 - 07:57

Imitation Learning or learning from expert trajectories can be implemented two different ways:

- Behavioral Cloning
 - a supervised learning problem that maps state/action pairs to policy
 - requires a large number of expert trajectories
 - copies actions exactly, even if they are of no importance to the task
- Inverse RL
 - learns the reward function from expert trajectories, then derives the optimal policy
 - expensive to run
 - indirectly learns optimal policy from the reward function

GAIL

GAIL is not exactly Inverse Reinforcement Learning because it's learns the policy, not the reward function, directly from the data. Yet, it's better than Behavioral Cloning and sometimes better than the experts, because it's doing Reinforcement Learning and it's not constrained to always be close to the expert.

GAIL, similar to a Generative Adversarial Networks, is composed of two neural networks. The Policy (Generator) network π_θ is trained using TRPO and the discriminator network D is a supervised learning problem trained with an ADAM gradient step on expert trajectories. Both networks have two hidden layers of 100 units each with a tanh activation.

The goal is to find a policy π_θ such that the discriminator cannot distinguish between states following the π_θ as opposed to those from π_{expert} .

Steps to train GAIL

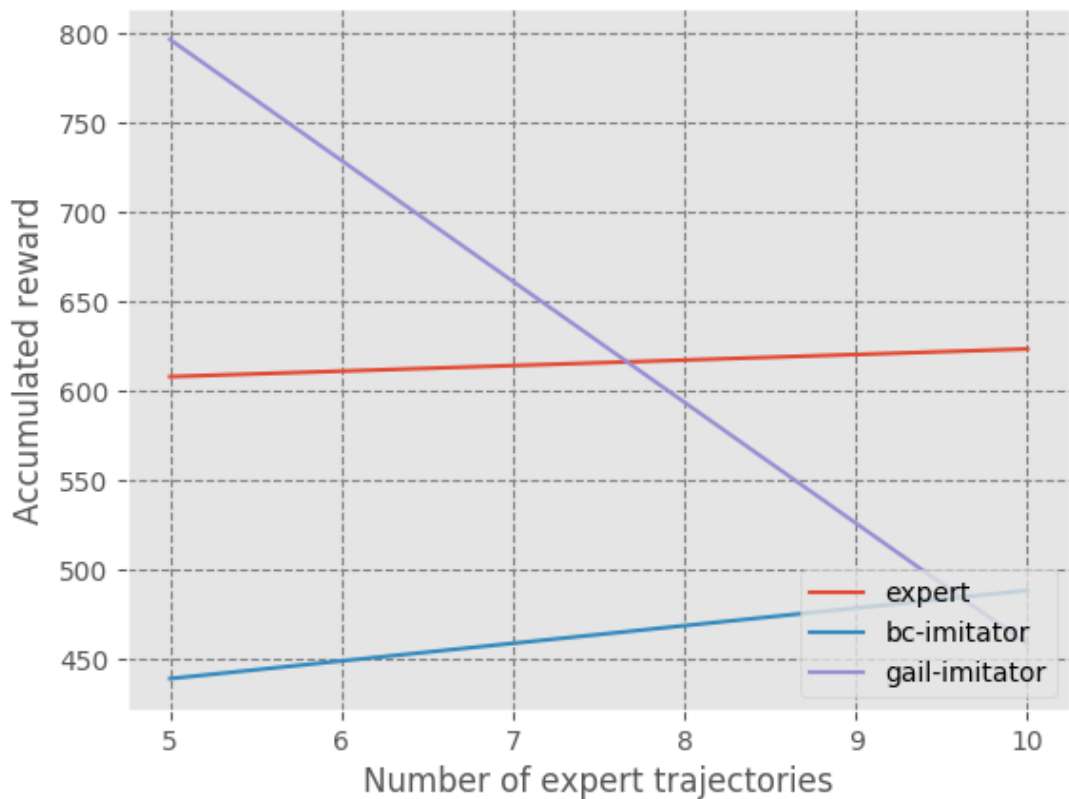
- Sample the expert trajectories
- Optimize the Policy π_θ
- Optimize the Discriminator D

I used OpenAI's Baseline GAIL code to train on MuJoCo: <https://github.com/hollygrimm/gail-mujoco>

Results

I ran GAIL and Behavioral Cloning on the following MuJoCo environments: Humanoid, HumanoidStandup, and Hopper. With five expert trajectories on Humanoid, GAIL was able to get better than "expert" results. These tests were run with only one seed, I would need to run many more seeds to make a conclusive statement.

Humanoid unnormalized scores



HumanoidStandup Trained on GAIL

0:00 / 0:14

References

1. Sergey Levine. "CS294 Inverse reinforcement learning". [Video](#) | [Slides](#)
2. Ng et al. "Algorithms for Inverse Reinforcement Learning". [PDF](#).
3. Fu et al. "Learning Robust Rewards with Adversarial Inverse Reinforcement Learning". [PDF](#).
4. Ziebart et al. "Maximum Entropy Inverse Reinforcement Learning". [PDF](#).
5. Wulfmeier et al. "Maximum Entropy Deep Inverse Reinforcement Learning". [PDF](#).
6. Finn et al. "Guided Cost Learning: Deep Inverse Optimal Control via Policy Optimization". [PDF](#).
7. Ho et al. "Generative Adversarial Imitation Learning". [PDF](#).

Tags

[Reinforcement Learning](#) [OpenAI](#) [GAIL](#)