Chanjoon Park
Feb 11, 2024

CS 285: Deep Reinforcement Learning, Decision Making, and Control
Assignment 1. Imitation Learning

3. Behavioral Cloning

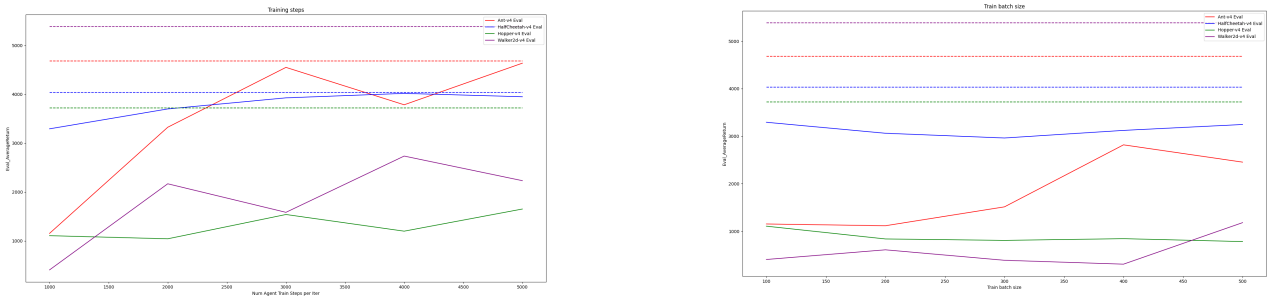
1. Run BC and report results

Gym Env	Avg. Return	Std. Return	Eval batch size	Episode	Expert's Avg. Return
Ant-v4	1153.34	142.75	5000	1000	4681.89
HalfCheetah-v4	3293.29	150.43	5000	1000	4034.80
Hopper-v4	1106.97	417.64	5000	1000	3717.51
Walker2d-v4	406.20	431.96	5000	1000	5383.31

Table 1: The mean and standard deviation of trained policy with given parameters.

Above results are collected from approximately 5 trajectories as you can see in Eval batch size and Episode and only the HalfCheetah-v4 environment shows at least 30% performance compared to the expert's policy.(Actually, 81.62% achieved.) The other environments show less than 30% performance, 24.63%, 29.78%, and 7.55% for Ant-v4, Hopper-v4, and Walker2d-v4, respectively.

2. Experiment with one set of hyperparameters.



(a) Training steps

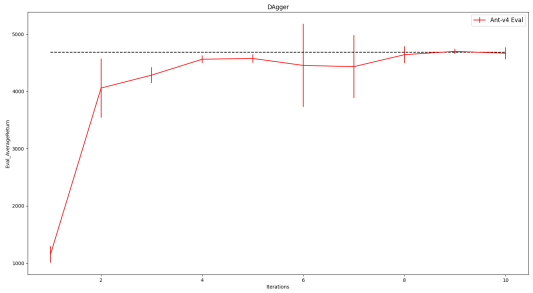
(b) Batch size

Figure 1: The performance of the trained policy with different hyperparameters. Red, blue, green and purple line represents the Ant-v4, HalfCheetah-v4, Hopper-v4, and Walker2d-v4 environments, respectively and dotted line represents the expert's policy.

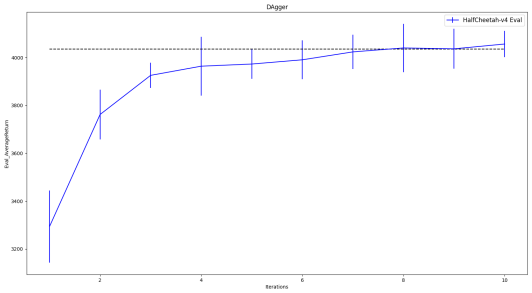
As we can see from the Figure 1a, the larger the training steps, the better the performance in general. However, the performance is not always improved as the training steps increase. For example, the performance of the Hopper-v4 and Walker2d-v4 environments is not improved as the training steps increase.

In the Figure 1b, there is no specific pattern that the performance is improved as the batch size increases.

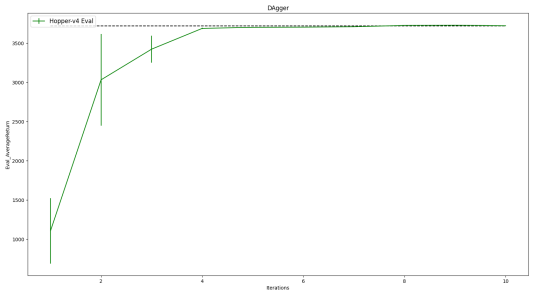
4. DAgger



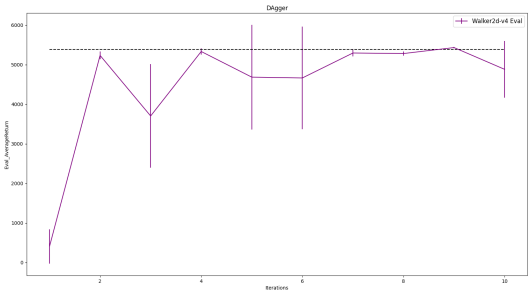
(a) And-v4.



(b) HalfCheetah-v4.



(c) Hopper-v4.



(d) Walker2d-v4.

Figure 2: The performance of the trained policy with DAgger for each environment. Black dotted line represents the expert's policy.

DAgger improves the performance of the trained policy over iterations. The performance of the trained policy is improved as the number of iterations increases generally.

Note *the commands to run the code are listed in the* README.md