Robust Learning from Observation with Model Misspecification

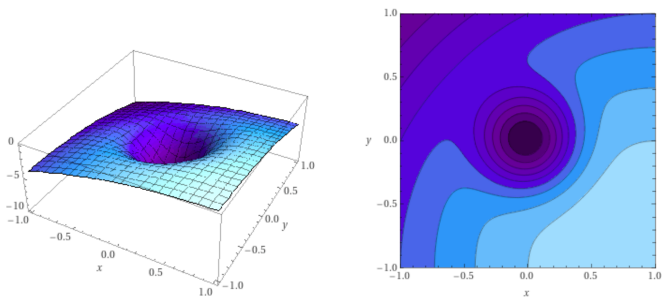
Anonymous Author(s)
Affiliation
Address
email

Keywords: Sim-to-real transfer, Imitation Learning, Learning from Observation

Dear reviewer,

here we present the results on our additional experiments evaluating the effectiveness of robust GAILFO for perturbations different than the ones induced by a friction or mass change in MuJoCo.

1 The environment



(a) Expert $\epsilon = 0.0$ (b) Expert $\epsilon = 0.05$

Figure 1: Schematics of the continuous gridworld environment

We consider a continuous gridworld, we denote the horizontal coordinate as x and vertical one as y. We enforce spatial constraints, i.e. $x \in [0, 1], y \in [0, 1]$.

The agent starts in coordinates $[0, 1]$ (the upper left corner) and the episode ends when the agent reaches the lower right region defined by the indicator function $\mathbf{1}\{x \in [0.95, 1], y \in [-1, -0.95]\}$.

The reward function is given by:

$$R(x, y) = -(x-1)^2 - (y+1)^2 - 80e^{-8(x^2+y^2)} + 10 \cdot \mathbf{1}\{x \in [0.95, 1], y \in [-1, -0.95]\} \quad (1)$$

In addition, the actions space for the agent is given by $\mathcal{A} = [-0.5, 0.5]^2$ and the transition dynamics are given by:

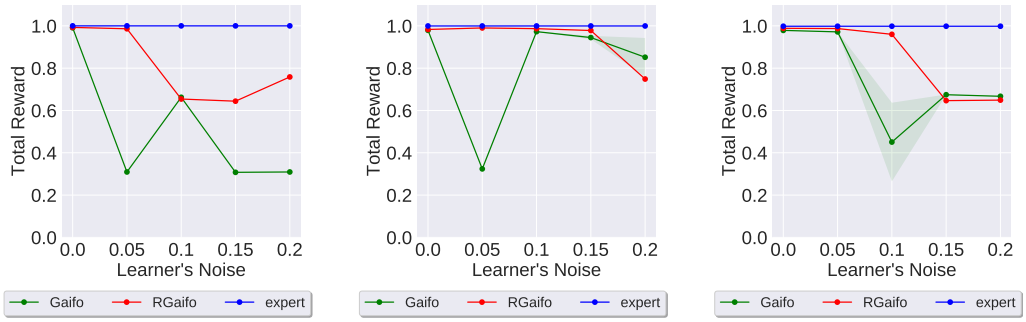
$$s_{t+1} = \begin{cases} s_t + \frac{a_t}{10} & \text{w.p.} \quad 1-\epsilon \\ s_t - \frac{s_t}{10||s_t||_2} & \text{w.p.} \quad \epsilon \end{cases} \quad (2)$$

The parameter ϵ can be varied to create a dynamic mismatch.

1.1 Experiments

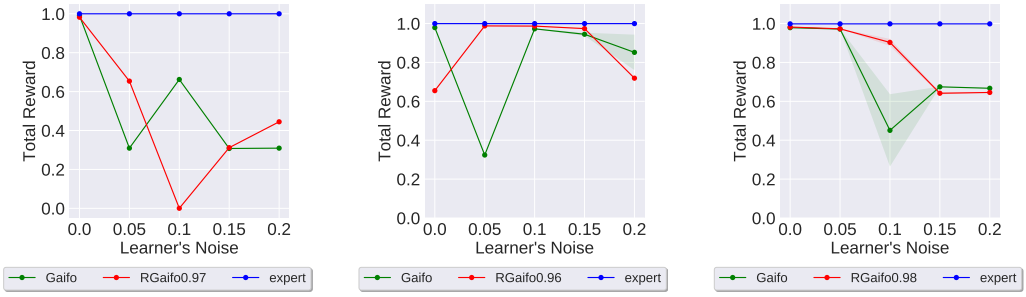
We use three experts trained with $\epsilon = 0.0, \epsilon = 0.05$ and $\epsilon = 0.1$.

Submitted to the 5th Conference on Robot Learning (CoRL 2021). Do not distribute.



(a) Expert $\epsilon = 0.0$ (b) Expert $\epsilon = 0.05$ (c) Expert $\epsilon = 0.1$

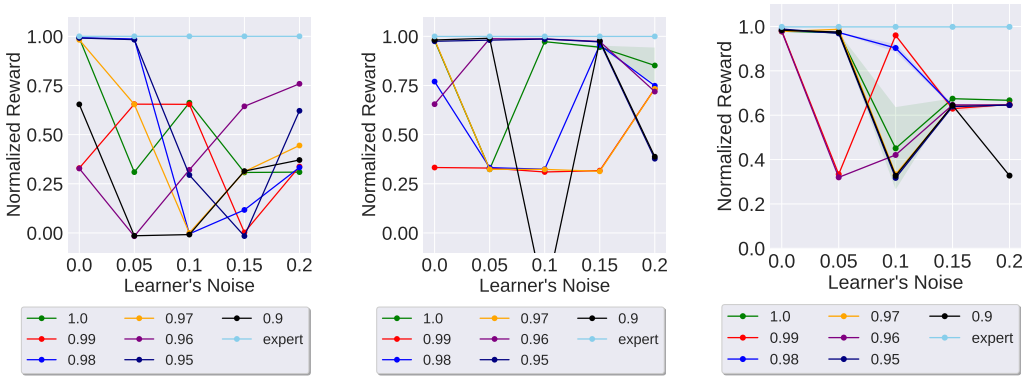
Figure 2: Comparison between Gailfo and Robust GailFO selecting the best α differently for each Learner Noise



(a) Expert $\epsilon = 0.0$ (b) Expert $\epsilon = 0.05$ (c) Expert $\epsilon = 0.1$

Figure 3: Comparison between Gailfo and Robust GailFO for the best selected α

The learners act in a different environment with the following values for ϵ: $0.0, 0.05, 0.1, 0.150.2$. In the following we report first the results of the comparison between GailFO and Robust GailFO with the best selected α and secondly the ablation study to assess the sensitivity to the parameter α.



(a) Expert $\epsilon = 0.0$ (b) Expert $\epsilon = 0.05$ (c) Expert $\epsilon = 0.1$

Figure 4: Ablation study for α