Imitation Learning for Latent Factors in Collaborative Multi-Agent Systems

Sharmistha Swasti Gupta†, Dana Hughes‡, Katia Sycara‡

†Department of Electronics and Communication Engineering, Indraprastha Institute of Information Technology, Delhi
‡Robotics Institute, School of Computer Science, Carnegie Mellon University

Introduction and Motivation

- Human behaviours can vary due to factors such as level of expertise, preference for a strategy etc.
- This introduces **unaccounted latent factors** for an agent trying to learn from a human and essentially yields **multiple distinct experts**.
- Currently, there does not exist a framework that enables multiple agents to take into account these latent factors and collaborate effectively with each other.

Background

- The GAIL framework [1] allows for directly recovering an expert policy from demonstration data.
  Objective Function:

$$\min_{\pi}\max_{D} \quad E_{\pi}\Big[logD(s,a)\Big] + E_{\pi_E}\Big[log(1-D(s,a))\Big] - \lambda H(\pi)$$

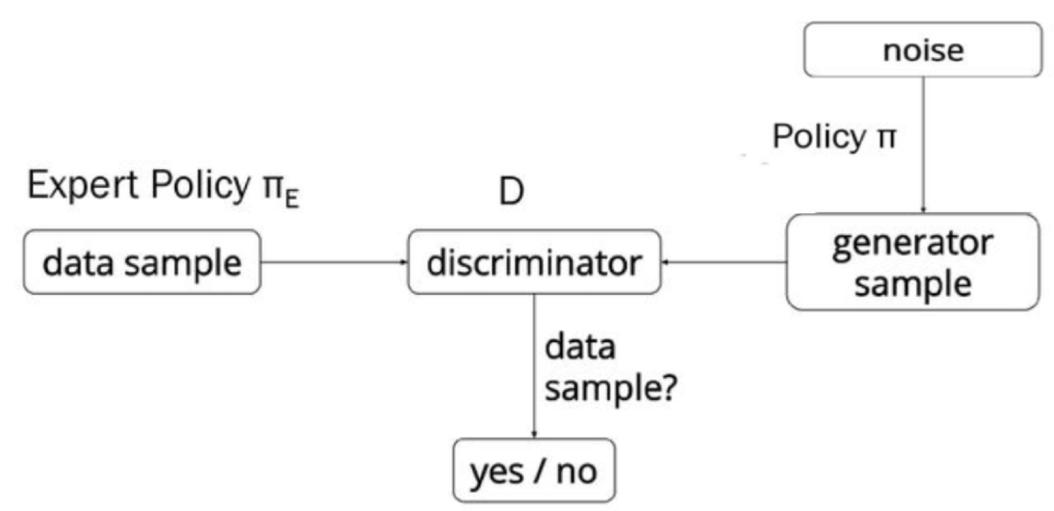

Fig. 1: Generative Adversarial Imitation Learning [1]

- Data from human expert demonstrations show significant variability due to the presence of latent factors that result in multiple distinct expert policies. **The InfoGAIL framework** [2] **infers the latency of such demonstrations** by learning a policy by introducing a latent variable in the policy function.
  Objective Function:

$$\min_{\pi,Q}\max_{D} \quad E_{\pi}\Big[logD(s,a)\Big] + E_{\pi_E}\Big[log(1-D(s,a))\Big] - \lambda_1 L_I(\pi(c),Q) - \lambda_2 H(\pi)$$

- **The MAGAIL framework** [3] **extends GAIL to multiple agents.** The cost function of MAGAIL is essentially the sum of the cost functions of all the agents.
  Objective Function:

$$\min_{\theta}\max_{\omega} \quad E_{\pi_\theta}\Big[\sum_{i=1}^{N} logD_{\omega_i}(s,a_i)\Big] + E_{\pi_E}\Big[\sum_{i=1}^{N} log(1-D_{\omega_i}(s,a_i))\Big]$$

Proposed Approach

Through this work, **we propose a novel framework that aims to combine two existing adversarial imitation learning algorithms: InfoGAIL** [2] **and MAGAIL** [3].

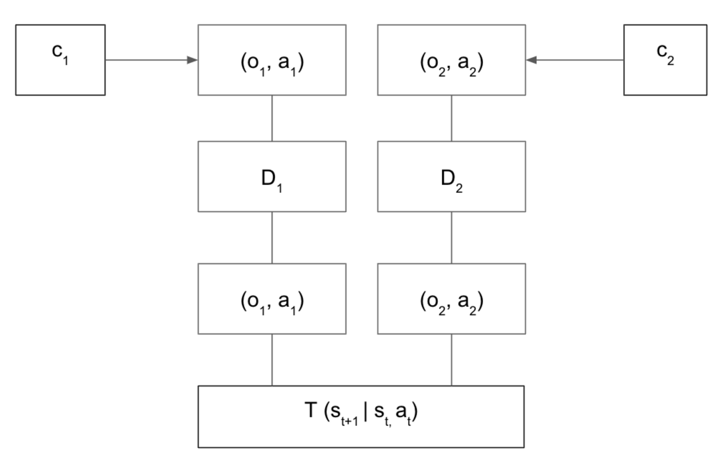

Fig. 2: Proposed Framework

Resultant objective function:

$$\min_{\theta,Q}\max_{\omega} \quad E_{\pi_\theta}\Big[\sum_{i=1}^{N} logD_{\omega_i}(s,a_i)\Big] + E_{\pi_E}\Big[\sum_{i=1}^{N} log(1-D_{\omega_i}(s,a_i))\Big]$$
$$- \lambda_1 L_I(\pi(c_1),Q) - \lambda_2 H(\pi(c_1)) - \lambda_3 L_I(\pi(c_2),Q) - \lambda_4 H(\pi(c_2))$$

Environment Setup

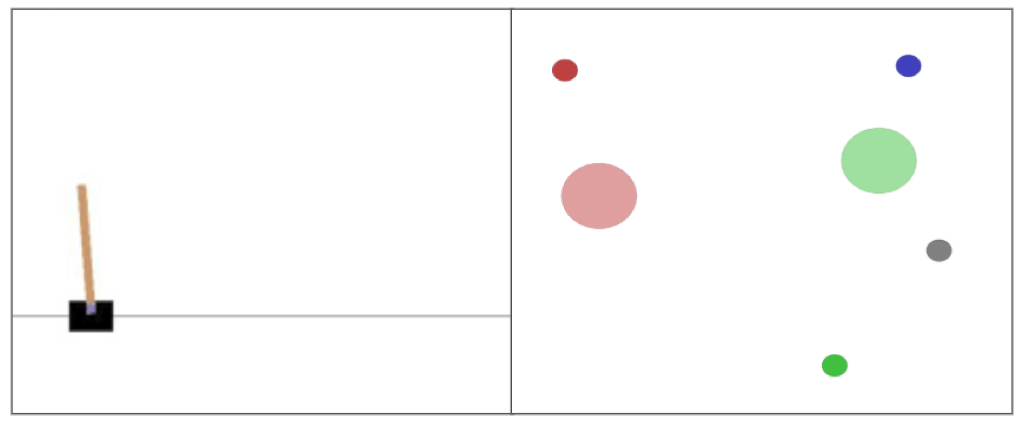

Fig. 3: (Left) Biased CartPole Environment: The block has a preferred position (along the x-axis) (Right) 4-Coloured Tasks Environment: The agents have preferred landmarks

- Biased CartPole Environment (Single Agent)
  - We introduce a **bias by specifying a desired position** of the block. The reward is then calculated as per the following Gaussian distribution (normalised to 1).

$$P(x) = \frac{1}{\sigma}e^{-(x-\mu)^2/2\sigma^2}$$

- 4-Coloured Tasks Environment (Multiple Agents)
  - Differently coloured agents and landmarks.
  - The collaborative task is for the two agents to cover the preferred landmarks.
  - **Agent preference is accounted for using colour** as a latent variable.

Results and Ongoing Work

- For the biased CartPole environment, we generated multiple experts (standard deviation is taken as 0.0001 in all cases) using PPO (Proximal Policy Optimization).

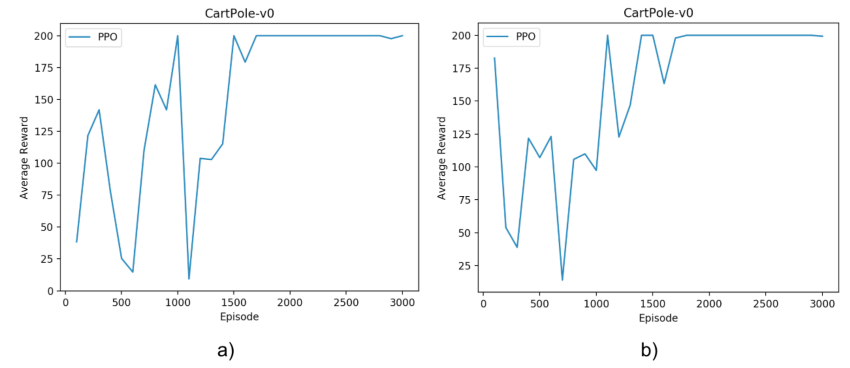

Fig. 4: Reward Function for Biased CartPole. Desired Position a) -2 b) +2

- For the 4-Coloured Tasks environment, we are using MADDPG (Multi-Agent Deep Deterministic Policy Gradient) to generate multiple experts such that the trained agent is able to continuously adapt to the preferences of the other agent.

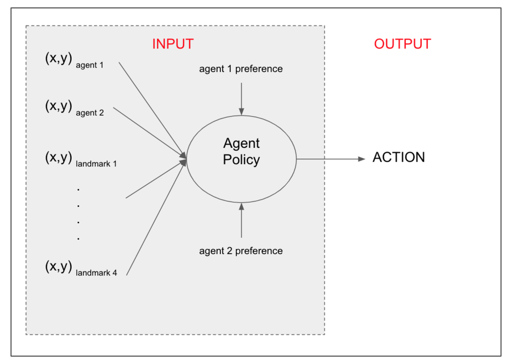

Fig. 5: Agent Policy

Future Work

- We will implement the proposed framework on our environments.
- Final experiments to be done in a Team Space Fortress environment.

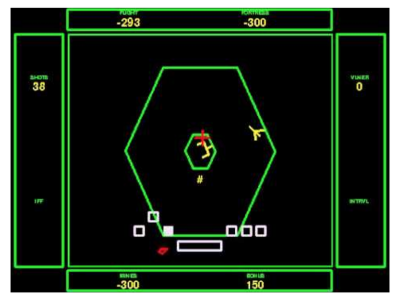

Fig. 6: Team Space Fortress Environment

Acknowledgements

The authors would like to thank members of the Advanced Agent-Robotics Technology Lab, especially Swaminathan Gurumurthy and Sumit Kumar, for providing valuable input. They would also like to thank Rachel Burcin, Prof. Dolan and the entire Robotics Institute Summer Scholars team.

References

[1] *Generative Adversarial Imitation Learning.* `https://slideplayer.com/slide/12976857/`. Accessed: 2019-08-12.

[2] Yunzhu Li, Jiaming Song, and Stefano Ermon. "Inferring The Latent Structure of Human Decision-Making from Raw Visual Inputs". In: *CoRR* abs/1703.08840 (2017). arXiv: `1703.08840`. URL: `http://arxiv.org/abs/1703.08840`.

[3] Jiaming Song et al. "Multi-Agent Generative Adversarial Imitation Learning". In: *CoRR* abs/1807.09936 (2018). arXiv: `1807.09936`. URL: `http://arxiv.org/abs/1807.09936`.