Dual Representation for Human-in-the-Loop Robot Learning

Stanford University

Dhruva Bansal, Yilun Hao, Ayano Hiranaka,
Roberto Martin-Martin, Chen Wang, Ruohan Zhang

Background

How can we make RL work on real robots?

- Incorporating human guidance can speed up learning
- Existing approaches face sample efficiency challenges

How do representations used by humans and robots during training differ?

- Humans: abstract, symbolic representation of other agents
- Robots: states and actions at millimeter, millisecond scale

Problem

Dual Representation Framework

- Fine grained state & action space for robots to perform control tasks
- Abstract, high-level representation for human to evaluate and guide robot (scene graph)

Evaluative Feedback

Human trainer monitor learning process and provide scalar feedback. Agent learns a policy that maximize human evaluation.

- Input: RL agent rollout trajectories, real-time human evaluations (-1, 0, +1)
- Output: trained RL policy

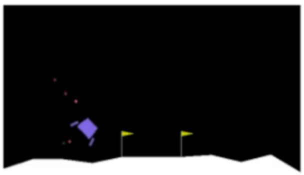
Preference Learning

Human trainer provide preference for set of pre-generated trajectories. Agent learns a reward function from human preference through IRL.

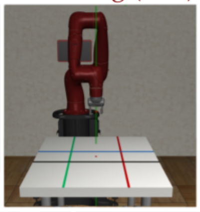
- Input: Randomly generated trajectories, human preference (0, 1)
- Output: reward weight

Experiments

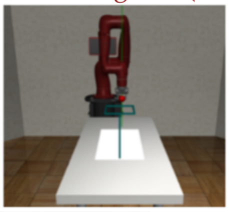
1: Luna-Lander

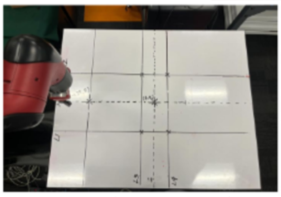


2: Reaching (Sim)

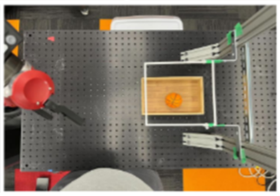


3: Placing Ball (Sim)





4: Reaching (Real)



5: Placing Ball (Real)

Methods

Baseline Models

Soft Actor-Critic (SAC)
Pure RL with no human feedback

TAMER+RL-100
Asking for feedback at every time step

TAMER+RL-50
Asking for feedback 50% of the time (uniform, random distribution)

TAMER+RL-25
Asking for feedback 25% of the time (uniform, random distribution)

Our Evaluative Feedback Model

Agent decides when to ask for human feedback according to change in abstract state (scene graph)

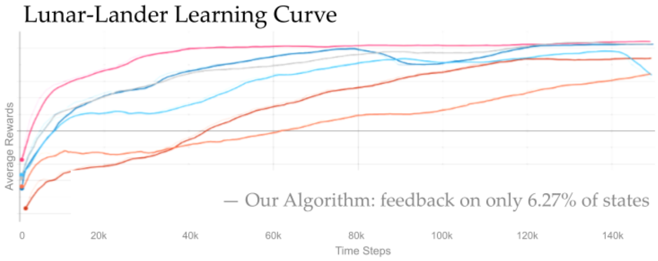
Our Preference Learning Model

Agent decides what query to ask human according to change in abstract state (scene graph)

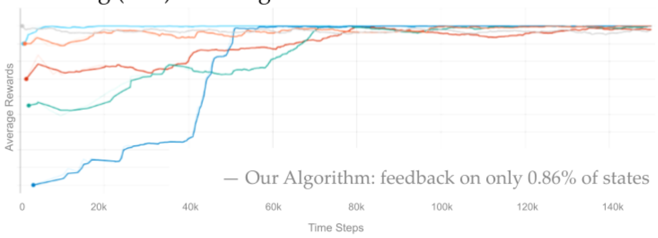
Results

Evaluative Feedback

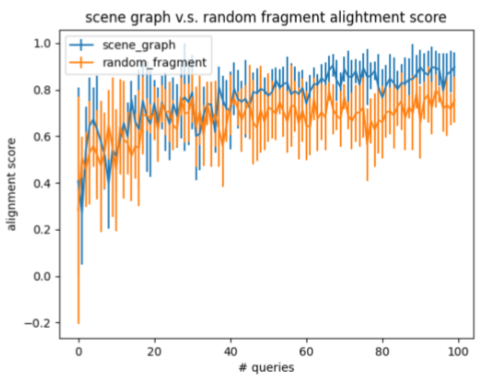
Lunar-Lander Learning Curve



Reaching (Sim) Learning Curve



Preference Learning



Conclusion

Our proposed learning algorithms based on the dual representation hypothesis can lead to significant improvements in task performance and sample efficiency