Coordinating and Incorporating Trust in Teams of Humans and Robots with Multi-Robot Reinforcement Learning

Christopher Amato & Stacy Marsella

Northeastern University

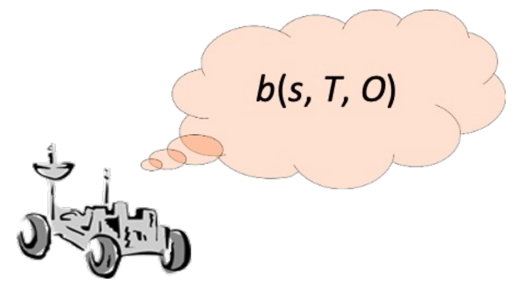
Project overview

Goal: How can teams of robots learn to collaborate with humans given the partial observability and uncertainty in human interactions as well as the vast differences in reasoning between robots and humans?

We plan to develop:

1. Teams of robots **learning** to assist humans even with **incorrect and incomplete human models**
2. Teams of robots learning to **coordinate and interact with humans using shared mental models**
3. Teams of robots learning to interact with humans by **incorporating trust**
4. **Test in simulation and hardware** in multiple scenarios

Modeling and Learning Human Models

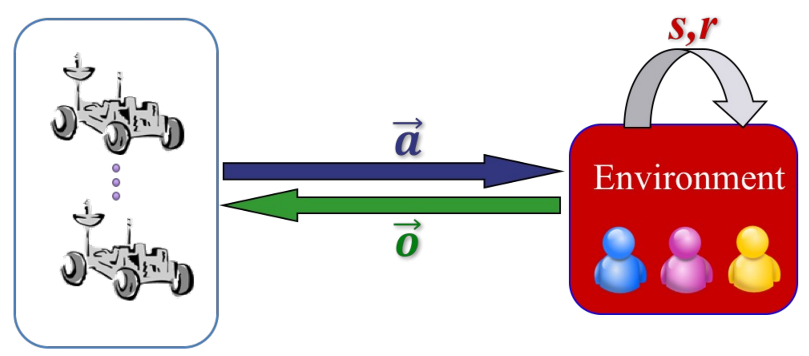


Overview

- Generate initial POMDP-based human models from simulation data
- Use these models in model-based (partially observable) reinforcement learning (RL)
- Improve models and solutions using Bayesian RL

POMDPs

- S, a set of states
- A, a set of actions
- T, the state transition model: $\Pr(s'|s,a)$
- R, the reward model: $R(s,a)$
- O, a set of observations
- Z, the observation model: $\Pr(o|s',a)$

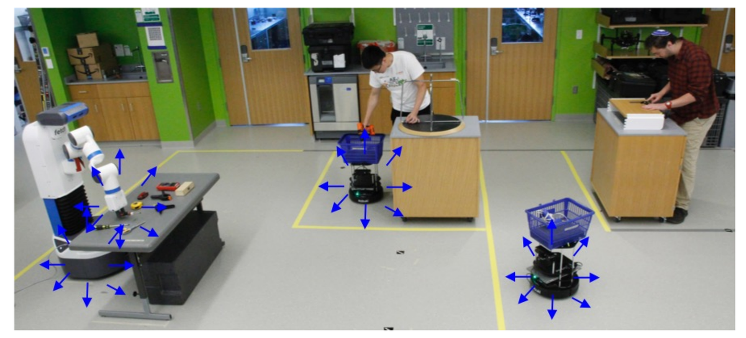


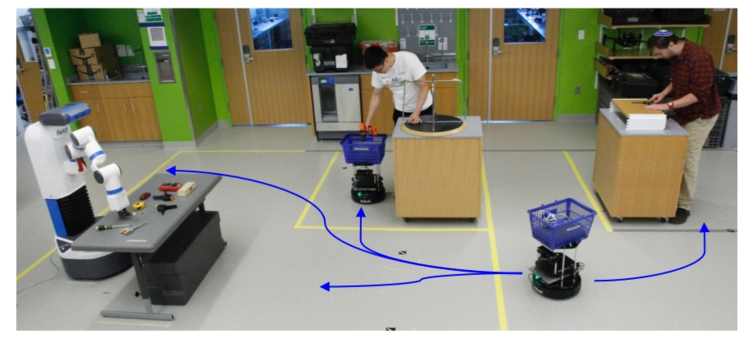
Generating initial model

- Will explore methods such as Automated Cognitive Behavior Analysis for generating from data with humans as part of the POMDP environment (based on [4])

Shared Mental Models

- Provide humans and robots with a shared model of the task as well as models of each other [6] by which to interpret and coordinate behavior
- Uses a hierarchical model over shared key coordination points
- Allows flexibility the low level but coordination and communication at the shared high level
- Extend Bayesian RL methods above to incorporate hierarchies
- Will be more scalable since they learn over fewer, higher-level actions and enable consistent mental models across the robots and people along with natural communication about those mental models.





Incorporating Trust and Interpretability

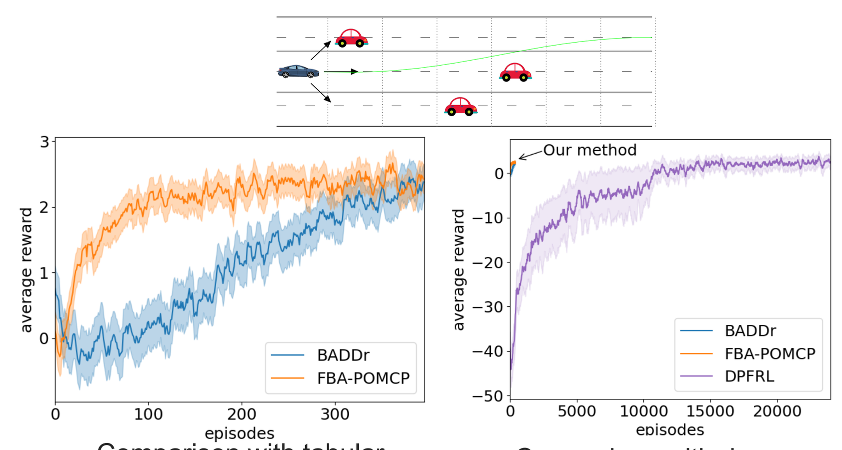
- True cooperation requires humans and robots to see each other as trusted teammates
- Therefore, we will incorporate trust into our POMDP models of human interaction
- And develop methods for improving trust by generating and sharing more interpretable robot solutions

Trust: Want to allow human teammates to appropriately judge the trustworthiness of robots using approaches based on the human response data and using Appraisal Theory to inform the design of trustworthy interaction and communication [5]

Interpretability: Interpretability in RL for HRI is important but general approaches don't currently exist. We will develop high-level interpretable representations of our Bayesian RL methods and other solutions along with uncertainty estimates.

Efficient Bayesian RL for POMDPs [1]

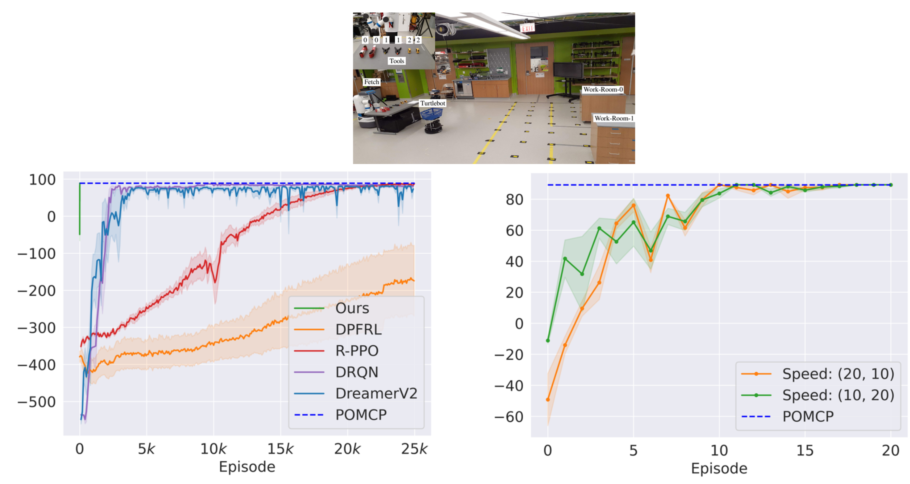
- Bayesian RL can optimally balance exploration and exploitation
- Ideal for online learning—optimally sample efficient so learn from very few interactions
- Computationally challenging, but our work combines deep and Bayesian RL to improve scalability [1]
- Bayes-adaptive deep dropout RL (**BADDr**), approximates the possible POMDP models with neural networks and approximates the Bayesian update with dropout in an ensemble of models
  - Slower than tabular Bayesian RL in small domains
  - But tabular Bayesian RL doesn't scale
  - Much faster than deep RL



Comparison with tabular Comparison with deep

On-Robot Bayesian Reinforcement Learning for POMDPs [2]

- BADDr will still require too many interactions with people to be practical in HRI settings
- Often little uncertainty about robots but significant uncertainty about people---exploiting structure in human interaction domains
- Developed a factored approach that reasons about each separately to (very) efficiently learn
- Tested on tasks where ordering of tools or speed of people changes
- Can learn optimal solutions in 10 episodes (or less)



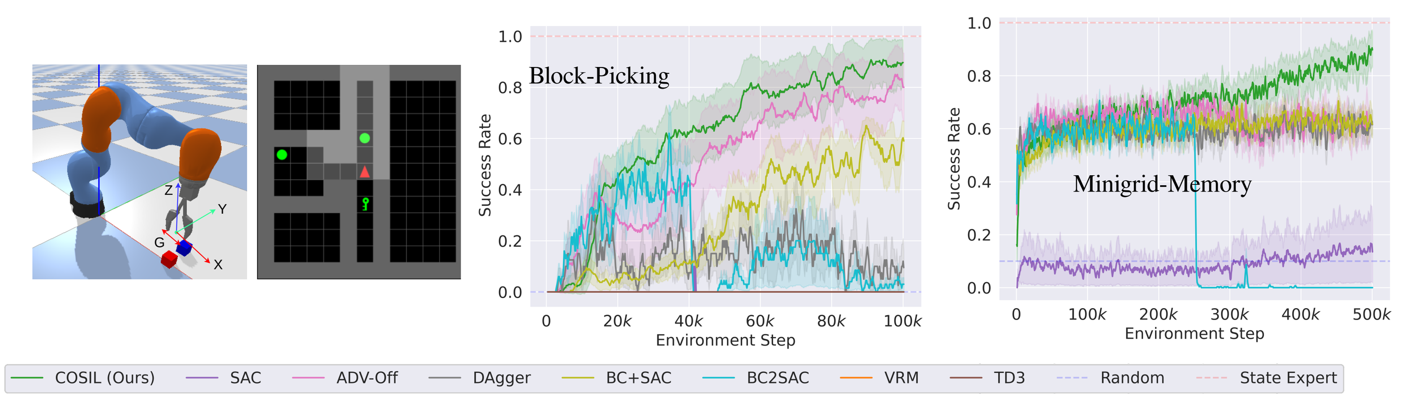
Speed task in sim Speed task in hardware

Leveraging Fully Observable Policies for Learning under Partial Observability [3]

- Also improving performance in non-Bayesian partially observable RL
- Here, use a MDP-based expert to help
- But an MDP won't do information gathering and could perform worse
- We maximize RL return while minimizing divergence between POMDP and MDP policies
- Performs better than other methods

$$J_\pi = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \left(R(s_t,a_t) - \alpha D(\mu(s_t), \pi(h_t))\right)\right]$$

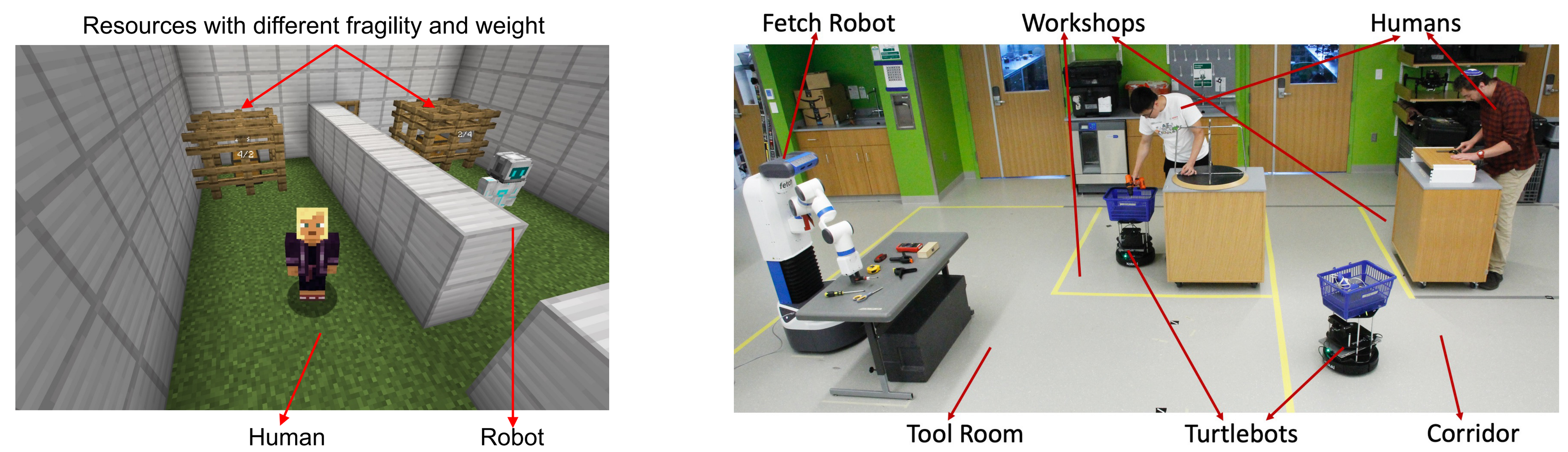
Where D is e.g., KL divergence



Experimental domains

Evaluate our methods in simulated scenarios and then hardware

- Developed two simulated scenarios to facilitate data gathering and testing methods
  - Coordinated search task in Minecraft and coordinated resource gathering task on the web
  - Now exploring trade-offs between task complexity, mental model reasoning and performance over time
- Will also test ideas in standard partially observable benchmarks
- Test in hardware using simple manufacturing setting (below) as well as more complex (simulated) search and rescue domain with humans teaming with aerial and ground robots



Current domains that will be used for testing (Minecraft and manufacturing)

[1] **BADDr: Bayes-adaptive Deep Dropout RL for POMDPs.** Sammie Katt, Frans A. Oliehoek and Christopher Amato. *In the Proceedings of the Eighteenth International Conference on Autonomous Agents and Multi-Agent System* (AAMAS-22), May 2022.
[2] **On-Robot Bayesian Reinforcement Learning for POMDPs.** Christopher Amato, Sammie Katt and Hai Nguyen. *Under submission*
[3] **Leveraging Fully Observable Policies for Learning under Partial Observability.** Hai Nguyen, Andrea Baisero, Dian Wang, Christopher Amato, Robert Platt. In *Proceedings of the 2022 Conference on Robot Learning* (CoRL-22), December 2022.
[4] **Analyzing Human Negotiation using Automated Cognitive Behavior Analysis: The Effect of Personality.** Pedro Sequeira and Stacy Marsella. In *Proceedings of Cognitive Science*, 2018.
[5] **Operationalizing Theories of Theory of Mind: A Survey.** Nik Gurney, Stacy Marsella, Volkan Ustun, David Pynadath. Operationalizing Theories of Theory of Mind: A Survey. *Springer, AAAI-FSS 2021.*
[6] **Effectiveness of Teamwork-Level Interventions through Decision-Theoretic Reasoning in a Minecraft Search-and-Rescue Task**, Pynadath et al.,, . Extended Abstract, *Proc. of 19th International Conference on Autonomous Agents and Multi-Agent System* (AAMAS-23), May 20223.

Award #1734497