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

Chris Amato and Stacy Marsella, Northeastern University

Challenge

- How can teams of robots learn to collaborate with humans given the uncertainty and vast differences in reasoning between robots and humans?

Fetch Robot · Workshops · Humans · Tool Room · Turtlebots · Corridor

General Plan

1. Teams of robots learning to assist humans even with incorrect and incomplete human models
   - Initial HRI POMDP models and then Bayesian deep reinforcement learning

2. … using shared mental models
   - For better communication and tight interaction

3. … incorporating trust
   - With human trust models and interpretability
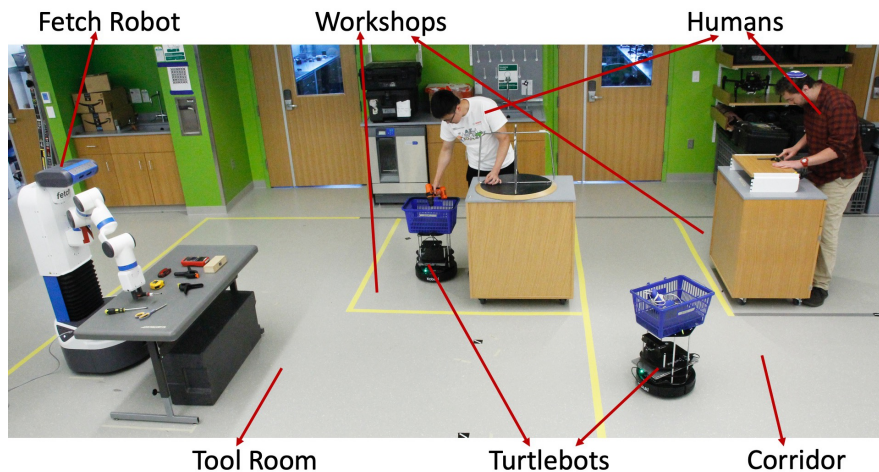
NRI: FND: Coordinating and Incorporating Trust in Teams of Humans and Robots with Multi-Robot Reinforcement Learning

Chris Amato and Stacy Marsella, Northeastern University

Efficient and safe reinforcement learning (RL)

- Sample efficient Bayesian deep RL for partial observable domains--- learns in hundreds of episodes rather than 10s of thousands (AAMAS 22)

- Safe multi-agent reinforcement learning during training and execution--- avoids collisions while still learning a high-quality solution (AAMAS 21)
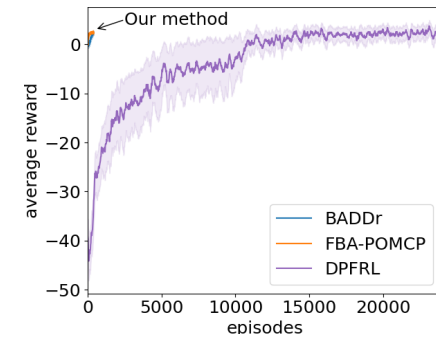
Current and future work

- Created simulation of a cooperative resource gathering task to support data collection at scale during pandemic

- Will also use Minecraft search and rescue task as a testbed

- Begun modeling trust and mental models of the other agents

- Also testing Bayesian deep RL methods in hardware (with simulated humans)

AAMAS 22

	MADDPG	MADDPG with Shield
Cross	207.20	0.00
Antipodal	14,419.20	0.00

Number of collisions (AAMAS 21)