Intent Prediction in Robotic Teleoperation Tasks Using Inverse Reinforcement Learning

Ben Weinstein-Raun, Laura Herlant; Advisor: Siddhartha Srinivasa

Motivation
To work toward a lower abandonment rate, how do we make assistive robotic arms easier to control for day-to-day use?

Task Representation
- A Markov Decision Process (MDP) consists of
  \[\langle U, A(s), T(s,a), R(s)\rangle\]
- We model the user as an agent in an MDP, where achieving some goal has a high reward
- An agent has a policy, \(\pi\)
- In reinforcement learning, we look for an optimal policy given a reward function

Assistance
- When the arm has more degrees of freedom than the input, modes allow the user to choose which ones to control
- Can we apply IRL to this?
  - We might accidentally learn something about the interface instead of the task
  - All observed trajectories control x-y-z and rotation separately
  - To explore possibilities, we’re currently gathering trajectory data

Testing Apparatus
- We set up targets at arbitrary positions in a 3x3x3 grid
  - The user pushes a button with the arm
  - Non-modal dataset: All buttons have the same orientation, and the gripper only moves in x-y-z
  - Modal dataset: Buttons are aligned with one x-y plane and have varying orientations; the user must switch modes to press them
- Collecting these datasets separately will allow us to more precisely determine users’ optimal policies
- Possible factors in cost function: mode switching, maximum speed, visibility of the path

Inverse Reinforcement Learning
- We have some “expert” trajectories (recorded by users) and we want to learn a cost function
- We have a hypothesis set of reward functions
- We find the one that maximizes the probability of the observed trajectories

References