RSS Paper

Active Preference-Based Learning of Reward Functions

Nikita Jaipuria

Aerospace Controls Laboratory
Department of Mechanical Engineering
Massachusetts Institute of Technology

August 11, 2017

Overview

- Objective:
  - model a **human's preference** for how a dynamical system should act
  - learn $R_H(x^0, \mathbf{u}_R, \mathbf{u}_H) = \sum_{t=0}^{N} r_H(x^t, u_R^t, u_H^t) = \sum_{t=0}^{N} \mathbf{w}^T \phi(x^t, u_R^t, u_H^t)$
- Problem Domain:
  - difficult to provide demonstrations of **desired** system trajectory (IRL)
  - assign numerical reward to an action/trajectory
- Main Idea: active preference-based learning
  - build on label ranking; learn from preferences/comparisons (**preference-based**)
  - system decides on what preference queries to make (**active**)
- Challenges/Contribution
  - complexity and continous nature of **queries**
  - **active synthesis** of queries satisfying system dynamics
  - **maximize volume removed** from continuous hypothesis space by each query

Algorithm

- Inputs: $\phi, N, f_{HR}, iter$
- Output: $p(\mathbf{w})$
- Step 1: Initialize $p(\mathbf{w}) \sim Uniform(B)$
- Step 1: **synthesize query** to remove as much volume as possible from the space of possible rewards (*constrained optimization*)

Slide Title 3

Slide Title 4

Questions

Questions?

Backup Slide 1

- Blah blah blah ...

References I