Human policy learning and interaction planning

Vaibhav Unhelkar1,2, Shen Li3, Julie Shah4
1 Interactive Robotics Group (IRG) 2 Rice University 3 CSAIL.

Agent Markov Model (AMM)

**Motivation**
- Modeling human policy is necessary
- Stochastic & multi-modal
- Manual specification
- Expensive & incomplete
- ⇒ Data-driven approach

**Key insights**
- Auto recover latent states
  - e.g., preference, fatigue, workload
  - Bayesian nonparametric
- Semi-supervise learning
  - Utilize partial specifications
  - e.g., dynamics, policy, change pts
  - Constrained variational inference

**Results**
- line world, highway
- Policy alignment \((T_x, \pi, b_x): \text{AMM} > \text{MaxEnt}\)
- KL weighted by relative counts of input data
- Aux inputs improve model alignment
- Aux inputs improve state decoding, prediction errors
- Normalized Ham dist bw inferred, true state seq


AdaCoRL: AMM + POMDP

**Motivation**
- Decision making under uncertainty
- Fluent & efficient HRI
- Sequential tasks with known objectives

**Results**
- Sandwich making (sim)
  - Model alignment: \(\text{AMM} \approx \text{supervised}\)
  - Total rewards: AdaCoRL > supervised
  - # safety stops: AdaCoRL < supervised
- Handover (sim)
  - Total rewards: AdaCoRL > supervised
- Handover (sim & user study)
  - # handovers: AdaCoRL > supervised

**Key insight 1**
- AMM human model learning
  - Inputs = partial specifications
  - \(S = \text{observable state space}\)
  - \(T_s(s' | s, a) = \text{obs state transition}\)
  - \(A = \text{action space}\)
  - Outputs
    - \(X = \text{latent state space}\)
    - \(T(x' | s, a, x) = \text{latent transition}\)
    - \(\pi(a | s, x) = \text{policy}\)
    - \(b_x(x) = \text{belief of initial x}\)

**Key insight 2**
- POMDP robot planning
  - State = \((\text{belief on x)}, s, s_R, s_E)\)
  - \(s_E = \text{env / task state (prespecified)}\)
  - \(A = \text{ar}\)
  - Reward (prespecified)
  - \(\Omega = (s, s_R, s_E), O = \text{prespecified}\)

CommPlan: AMM + POMDP + Communication

**Motivation**
- Effective comm. is critical for fluent HRI
- Bidirectional comm.
- Deciding “if,” “when,” “what” to comm.

**Communication types**
- Inform: “I am going to do action at X.”
- Ask: “Where are you going?”
- Command: “Please make next sandwich at X.”

**Results**
- Sandwich making task
  - Shared reward: CommPlan > silent > hand-crafted
  - Task time: CommPlan < silent < hand-crafted
  - # robot comm.: CommPlan > hand-crafted > silent

Unhelkar, Vaibhav, Shen Li, and Julie Shah. "Decision-making for bidirectional communication in sequential human-robot collaborative tasks." HRI 2020