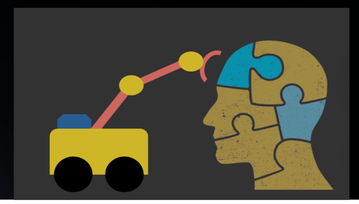
EAGER: Reconciling Model Discrepancies in Human-Robot Teams

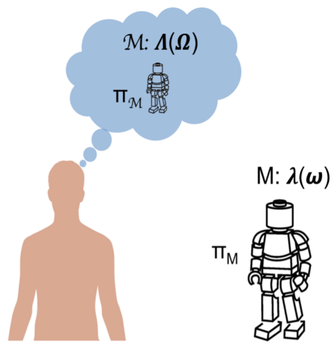
Yu ("Tony") Zhang, Arizona State University



1. Motivation

- Teammates have many conscious and subconscious expectations of others in terms of their plans or behaviors
- The expected domain model (EM, for generating expectations) and the true domain model (DM) may differ, leading to unmatched expectations, loss of situation awareness and trust

➢ This calls for generalized planning and learning methods for domain model reconciliation



2. Scientific Impact

- A step towards ubiquitous collaborative robots with non-expert users, such as for autonomous cars, household robotic assistants, etc.
- Address human biases towards robots to improve team situation awareness and trust

3. Technical Impact

- Generalize planning methods to open-world domains where both the true domain model and the human's model of expectation (hidden) are considered
- Generalize learning methods for situations with model discrepancy

4. Solutions and Progresses

- *Planning methods*
  - ✓ *AAMAS20: Explicable planning*: use plan distances to define a distance metric for generating explicable plans
  - ✓ *IROS20: Online explanation generation* for explaining model differences: use planning methods to break an explanation into multiple pieces and communicate them online to reduce cognitive load
  - ✓ *ICRA21 (accepted): Progressive explanation generation* for model differences: apply learning to determine an appropriate order for communicating different parts of an explanation
  - ✓ *HRI21 (accepted): Active explicable planning*

4. Solutions and Progresses

- *Learning methods*
  - ✓ *AAAI20: Generalized reward learning* under biased domain dynamics: apply IRL may lead to learning the opposite human preferences when humans are biased
  - ✓ *Under review*: *Domain concretization*: generalize planning methods to cases where the human's model is more complete
  - ✓ *In progress*: Generalize explicable planning to continuous domain with unknown human model of expectation
  - ✓ *In progress*: Consider the effectiveness of additional modalities for communicating model discrepancy

5. Broader Impact: Societal

- *Ubiquitous collaborative robots require robotic technologies that support human-robot teaming*
- *Safety and trust issues*
- *Co-bot technology for improving our everyday life; public awareness*
- *Interpretable and explainable AI (AI explains complex behaviors and their rationale)*
- *Synergy with other programs*

6. Broader Impact: Education

- *Keynote speaker at Intel, Chandler*
- *Invited talk at IROS 2020*
- *Judging for Intel ISEF, involving high school students from around the world*
- *Engineering projects for graduate student at ASU*
- *CSE 591 "Human-Aware Robotics", covering research methods developed in this research*

7. Broader Impact in numbers

- *4 Ph.D. students partially supported*
- *3 MS students, 1 graduated*
- *2 successful proposals, 2 others in development based on the results*
- *6 publications; more under review or in progress*
- *2 invited talks on the research topic*
- *1 graduate class on the research topic*

2021 NRI & FRR Principal Investigators' Meeting
March 10-12, 2021

Award ID#:1844524