Improving Human-Robot Interaction through Explainable Reinforcement Learning

Aaquib Tabrez
University Of Colorado Boulder
Boulder, CO 80309
mohd.tabrez@colorado.edu

Bradley Hayes
University Of Colorado Boulder
Boulder, CO 80309
bradley.hayes@colorado.edu

I. Introduction and Background

Gathering the most informative data from humans without overloading them remains an active research area in AI, and is closely coupled with the problems of determining how and when information should be communicated to others [12]. Current decision support systems (DSS) are still overly simple and static, and cannot adapt to changing environments we expect to deploy in modern systems [3], [4], [9], [11]. They are intrinsically limited in their ability to explain rationale versus merely listing their future behaviors, limiting a human's understanding of the system [2], [7]. Most probabilistic assessments of a task are conveyed after the task/skill is attempted rather than before [10], [14], [16]. This limits failure recovery and danger avoidance mechanisms. Existing work on predicting failures relies on sensors to accurately detect explicitly annotated and learned failure modes [13]. As such, important non-obvious pieces of information for assessing appropriate trust and/or course-of-action (COA) evaluation in collaborative scenarios can go overlooked, while irrelevant information may instead be provided that increases clutter and mental workload. Understanding how AI models arrive at specific decisions is a key principle of trust [8]. Therefore, it is critically important to develop new strategies for anticipating, communicating, and explaining justifications and rationale for AI driven behaviors via contextually appropriate semantics.

II. Current Work

To address the need for robots to effectively collaborate with humans, we have been working on methods for establishing a shared mental model amongst teammates. In the case of incongruous models, catastrophic failures may occur unless mitigating steps are taken. To identify and remedy these potential issues, we proposed a novel mechanism for enabling an autonomous system to detect model disparity between itself and a human collaborator, infer the source of the disagreement within the model, evaluate the potential consequences of this error, and finally, provide human-interpretable feedback to encourage model correction. This process effectively enables a robot to provide a human with a policy update based on perceived model disparity, reducing the likelihood of costly or dangerous failures during joint task execution.

We modelled our framework upon the assumption that sub-optimal collaborator behaviour is the result of a mis-informed understanding of the task rather than a problem with the collaborator's rationality. Formulated as a Markov Decision Process, a human's sub-optimal decision-making could be attributable to a malformed policy given an incorrect task model. Building on this assumption, we created an autonomous collaborative agent that was able to: 1) infer the most likely reward function used as a basis for a human's behavior; 2) identify the single most detrimental missing piece of the reward function; and 3) communicate this back to the human as actionable information to enable the collaborator to update their reward function (task comprehension) and policy (behavior).

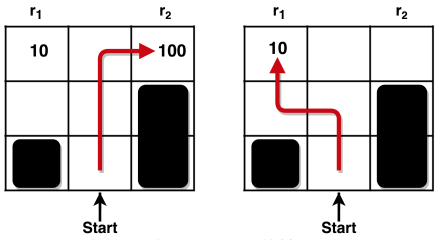

Fig. 1: Task executions given two different comprehensions of a gridworld domain's reward function. Left: Both rewards are known based on the path taken. Right: Reward r_2 is assumed to be unknown based on the path taken by the human.

A. A Framework for Reward Augmentation and Repair through Explanation

Our proposed framework, *Reward Augmentation and Repair through Explanation* (RARE), utilizes a Partially Observable Markov Decision Process (POMDP) coupled with a family of Hidden Markov Models (HMMs) to infer and correct a collaborator's task understanding during joint task execution through updates to their reward function. Once a plausible reward function is discovered that explains the collaborator's behaviour, a repairing explanation can be generated and provided if the benefit of correction outweighs the consequences of ignoring it and proceeding with the task.

1) Estimation of Reward Comprehension

To determine which components of the reward function the human collaborator is using, RARE utilizes an HMM that incorporates both state features of the world ("world features") and latent state features that indicate knowledge of the corresponding components of domain's reward function ("comprehension features").

2) Collaborative Task execution and Reward Repair

For a given collaborative task, we define the RARE agent's behaviors with a policy that solves an augmented POMDP (RARE-POMDP) which uses augmented states, consisting of both world features and comprehension features. RARE-POMDP's estimate of which reward function the collaborator is following is proportional to the likelihood that their behavior was informed by a policy derived from it (Figure 1). The RARE-POMDP introduces the opportunity for the agent to make the decision to execute social actions aimed at better informing a collaborator about the domain's reward function, in addition to traditional task-progressing actions. In other words, the agent may execute a communicative action to explicitly inform a collaborator about part of the reward function which is missing from their belief, directly changing the value of a latent comprehension feature.

3) Explanation Generation

The RARE framework allows an agent to estimate a collaborator's reward function during joint task execution. This is far more useful in a collaborative context when paired with actions that enable one to augment a collaborator's understanding of the task. For our application domain, we proposed an algorithm that autonomously produces statements capable of targeted manipulation of a collaborator's comprehension features based on anticipated task failures.

B. Results

To quantify the viability and effectiveness of RARE, and to evaluate our hypotheses within a live human-robot collaboration, we conducted a between-subjects user study using a color-based collaborative Sudoku variant and an autonomous robot (Rethink Robotics Sawyer). Study participants were assigned into one of two conditions: 1) *Control*: The robot interrupts users that are about to make mistakes, indicating to them that it will cause task failure; 2) *Justification*: The robot interrupts users about to make mistakes, indicating that it will cause task failure and explaining which game constraint will inevitably be violated.

Subjectively, we were able to confirm the following two hypotheses: 1) *H1*: Participants will find the robot more helpful and useful when it explains why a failure may occur and 2) *H2*: Participants will find the robot to be more intelligent when it gives justifications for its actions [15].

Objectively, we observed that there were more terminations (irremediable mistake) of the game during the control condition as compared to the justification condition $(80\%$ vs $20\%)$. From the responses, we were able to deduce reasons for more terminations in the control condition — *participants did not trust Sawyer when it indicated that the human was about to make a critical mistake, when it did so without further explanation*. They were skeptical with respect to Sawyer, who was not providing accompanying justification for its judgment of their move. We also found evidence in the surveys supporting the notion that *providing justification alongside feedback leads to a more positive user experience*. Our results highlight that justification is an important requirement for a robot's corrective explanation. Hence, we validated that our contribution is not a solution in search of a problem, but addresses an important, underexplored capability gap in the HRI and Explainable AI literature.

III. Future Work

One of the drawbacks of RARE is that our formulation of 'comprehension features' causes a combinatorial expansion of state space, with non-trivial reward functions causing RARE to easily become intractable. There are many potential approaches for addressing this problem in terms of accommodating arbitrary reward functions with a reduced set of comprehension features (i.e., making a priori assumptions about what one's collaborator knows), or providing more easily solved approximations of the true reward function (i.e., approximation using fewer reward factors).

State abstraction is a vital application for policy iteration, enabling effective generalization that provides dramatic complexity reduction in exchange for sacrificing fine state granularity [5]. Abel et al. provided state abstraction theory out of the traditional single task setting by exploring the benefits and pitfalls of learning and planning with various types of state abstractions [1]. As an extension to our current RARE framework, we will be using graph-theoretic techniques to investigate the conductivity of MDP subgraphs and discontinuities within the MDP's converged state-value function for contraction into an abstract states that go beyond the limitations of existing methods [6]. This module of state abstraction will help RARE enabled agents to tractably estimate human policies and provide policy updates by way of explanation even within complex tasks.

The RARE framework provides an explanation for the sub-optimal behavior of the human, but in its current form lacks a necessary comprehensibility of its optimal policy. Recall that an agent's policy is a scalar representation of that agents consideration of risks and rewards given a state and prospective action. The factors that went into that consideration are lost during the computation of an optimal policy, as future outcomes are merged together into an *expectation* during policy (or value function) updates. Therefore, reinforcement learners cannot articulate rationale for its actions or 'concerns' in a human interpretable way until this is addressed through novel bookkeeping techniques.

IV. Conclusion

In conclusion, we have proposed a novel framework for estimating and improving a collaborator's task comprehension and execution. By characterizing the problem of sub-optimal performance as evidence of a malformed reward function, we introduce mechanisms to both detect the root cause of the sub-optimal behaviour and provide feedback to the Human to repair their decision-making process. In the future, we would continue focusing on the intersection of explainable AI (xAI) and human-robot collaboration : 1) *Developing a scalable framework for human policy estimation and reward coaching* and, 2) *Propose a novel framework of policy explanation for the robot by bookkeeping during policy iteration.*

REFERENCES

[1] D. Abel, D. Arumugam, L. Lehnert, and M. Littman. State abstractions for lifelong reinforcement learning. In *International Conference on Machine Learning*, pages 10–19, 2018.

[2] M. Aitken, N. Ahmed, D. Lawrence, B. Argrow, and E. Frew. Assurances and machine self-confidence for enhanced trust in autonomous systems. In *RSS 2016 Workshop on Social Trust in Autonomous Systems*, 2016.

[3] D. Albers, M. Correll, and M. Gleicher. Task-driven evaluation of aggregation in time series visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 551–560. ACM, 2014.

[4] Y. Boussemart and M. Cummings. Behavioral recognition and prediction of an operator supervising multiple heterogeneous unmanned vehicles. *Humans operating unmanned systems*, 2008.

[5] N. Gopalan, M. desJardins, M. L. Littman, J. MacGlashan, S. Squire, S. Tellex, J. Winder, and L. L. Wong. Planning with abstract markov decision processes. In *ICAPS*, 2017.

[6] B. Hayes and B. Scassellati. Autonomously constructing hierarchical task networks for planning and human-robot collaboration. In *International Conference on Robotics and Automation (ICRA)*. IEEE, 2016.

[7] B. Hayes and J. A. Shah. Improving robot controller transparency through autonomous policy explanation. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 303–312. ACM, 2017.

[8] M. Hind, S. Mehta, A. Mojsilovic, R. Nair, K. N. Ramamurthy, A. Olteanu, and K. R. Varshney. Increasing trust in ai services through supplier's declarations of conformity. *arXiv preprint arXiv:1808.07261*, 2018.

[9] T. M. Howard, S. Tellex, and N. Roy. A natural language planner interface for mobile manipulators. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 6652–6659. IEEE, 2014.

[10] R. A. Knepper, S. Tellex, A. Li, N. Roy, and D. Rus. Recovering from failure by asking for help. *Autonomous Robots*, 39(3):347–362, 2015.

[11] H. Kress-Gazit, G. E. Fainekos, and G. J. Pappas. Temporal-logic-based reactive mission and motion planning. *IEEE transactions on robotics*, 25(6):1370–1381, 2009.

[12] K. G. Lore, N. Sweet, K. Kumar, N. Ahmed, and S. Sarkar. Deep value of information estimators for collaborative human-machine information gathering. In *Proceedings of the 7th International Conference on Cyber-Physical Systems*, page 3. IEEE Press, 2016.

[13] P. Pastor, M. Kalakrishnan, S. Chitta, E. Theodorou, and S. Schaal. Skill learning and task outcome prediction for manipulation. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3828–3834. IEEE, 2011.

[14] C. Plagemann, D. Fox, and W. Burgard. Efficient failure detection on mobile robots using particle filters with gaussian process proposals. In *IJCAI*, pages 2185–2190, 2007.

[15] A. Tabrez, S. Agrawal, and B. Hayes. Explanation-based reward coaching to improve human performance via reinforcement learning. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019.

[16] M. L. Visinsky, J. R. Cavallaro, and I. D. Walker. Robot fault detection and fault tolerance: a survey. *Reliability Engineering and System Safety*, 46(2):139–158, 1994.