Enabling Robot Teammates to Learn Latent States of Human Collaborators

Vaibhav Unhelkar, Charlie Guan, Nicholas Roy and Julie Shah
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
Email: {unhelkar, cguan, nickroy, julie_a_shah}@csail.mit.edu

Abstract

We are interested in designing collaborative robots that can seamlessly interact in complex domains - such as, collaborative manufacturing and disaster response - with human teammates. To be successful teammates, such robots need the ability to model, predict and adapt to their human collaborators. However, modeling the behavior of human teammates is challenging - since human decisions often depend on factors that are latent and difficult to specify [4]. In this extended abstract, we describe this challenge for modeling humans for designing robot collaborators, summarize our algorithmic solutions towards this problem and conclude with a description of a human-robot collaboration scenario designed to evaluate our approach.

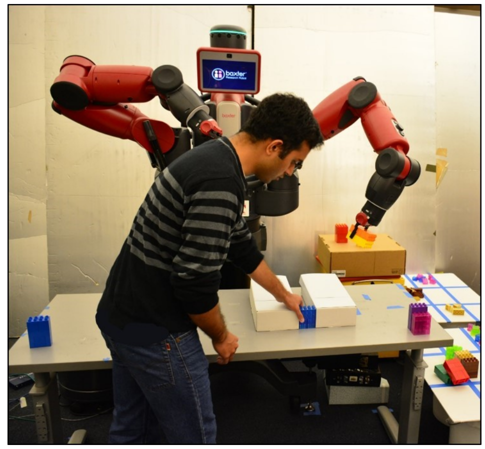

Fig. 1. A human-robot team performing a collaborative assembly task in a shared environment.

I. Introduction

As a prototypical scenario of human-robot collaboration, let us consider a human and a robot performing collaborative assembly in a shared environment as depicted in Fig.1. Both the human and the robot can manipulate objects in the shared environment to accomplish their sequential task. The presence of the human teammate renders the robot's environment both dynamic and unstructured. Prior studies of human-robot interaction have shown that anticipating, communicating and adapting to teammates is critical to achieving fluent human-robot interaction [6]. This can be achieved by enabling the robot to (i) model and predict human's actions, and (ii) plan its actions and communications using appropriate planning methods to accomplish the task [7, 11].

However, generating models of human decision-making that enable this anticipatory behavior can be challenging – a key reason being that human decisions may depend on factors (*states*) that are both difficult to specify and observe (i.e., latent). For instance, in our example scenario, the human's decisions (e.g., which object to pick, the path used to pick up an object) may depend on a variety of factors, such as, goal, preference, location of human and robot, human's trust in the robot, etc. Further, these factors could be both latent and dynamic.

A typical approach while modeling human's (or another agent's) decision-making is to first specify a set of factors (s) that impact the decisions (a), and then use a learning approach (either supervised learning, inverse reinforcement learning or active learning) to learn a policy (π, i.e., mapping from states to actions) describing human behavior using data or via interaction [1, 2, 12]. Recent research in human-robot interaction provides successful examples of employing such an approach to estimate human's reward and policy for robotics (e.g., [8]). However, a policy learnt by such an approach ($\hat{\pi}$) – irrespective of the available data – can only model the effect of pre-specified decision factors on human's decisions, i.e.,

$$\hat{\pi} = \hat{\pi}(a|s). \tag{1}$$

This learnt policy ($\hat{\pi}$), in general, will be an approximation of the true policy (π), which may additionally depend on unmodeled factors (x) of human decision-making, i.e.,

$$\begin{aligned} \pi &= \pi(a|s,x) \\ &\approx \hat{\pi}(a|s). \end{aligned} \tag{2}$$

Hence, to learn the effect of unmodeled factors on human's policy, we are developing approaches that infer presence of any unmodeled states (x) and their dynamics in addition to learning a policy (π).

II. Learning Decision-Making Models via Human-in-the-Loop Inference

We pose the problem of learning a human's sequential decision-making behavior without complete state specification (i.e., factors that impact decisions) as one of Bayesian inference. Specifically, inference of the following quantities is required to model decision-making,

- unknown number of unmodeled states, x,
- the dynamics of unmodeled states (denoted by transition function, $T_x(x'|x, s, a)$), and
- the state-dependent policy, $\pi(a|x, s)$.

Similar to prior approaches [12], we use execution traces – i.e., time series of observable, modeled states (s) and human's decisions (a) – as input to the algorithm. However, in contrast to prior approaches, we additionally seek to learn both the unmodeled states (x) and the variables (π, T_x) that depend on them.

This problem of learning the decision-making model, essentially, corresponds to that of joint clustering and segmentation of the observed execution traces (i.e., time series of (s, a)-pairs). In this analogy, the clusters correspond to the latent states, and segmentation corresponds to inferring the transition of these latent states. Since the number of these latent states (or clusters) is unknown, we adopt a Bayesian nonparametric approach to specify a generative model for decision-making. Briefly, we use hierarchical Dirichlet process (HDP) priors inspired by the HDP-HMM model of Teh et al. [9] to model the number of latent states and their transition function.

To perform Bayesian inference over our generative model of decision-making, we have developed a sampling algorithm that provides samples of number of latent states (x), transition function (T_x) and policy (π). Since multiple models may have similar posterior give the input data – identifying the true model among these sampled candidate models can be difficult. Hence, we adopt a *human-in-the-loop inference* approach that involves querying the human decision-maker to acquire additional information and guide the inference process.

In contrast to prior works that use querying to learn models [2, 3], learning unmodeled states necessitates the use of novel query types; namely, query types that are based on only the human and robot's *common ground* of observed states (s) and actions (a) but can help in the inference of unmodeled states (x) are needed. Since, querying a human typically has associated costs - we have developed an active querying approach that selects the most informative query, i.e., poses a query which minimizes the uncertainty over the inferred decision-making model. Through simulation experiments, we observe that querying improves model estimation performance. In our on-going work, we are exploring extensions that utilize function approximation and include tighter coupling between the sampling and querying process to accelerate the inference process using human input.

III. Implications for Robot Teammates

By algorithmically modeling the previously unmodeled states, the robot teammate has the potential to improve the predictions of its human collaborator and make more informed action and communication decisions during collaborative tasks. We intend to evaluate this potential benefit using the setup depicted in Fig. 1. Specifically, we consider scenarios inspired by collaborative assembly in confined spaces, where both the human and the Baxter robot work in a shared environment and have the ability to communicate with each other. In such scenarios, the human-robot team will have to perform sequential tasks that require manipulating objects in constrained spaces (e.g., peg-in-hole) while sharing their workspaces. To accomplish the robot's component of the collaborative task, we have developed an approach for near-optimal contact-rich manipulation which leverages environmental contact [5]. This enables human-safe robotic systems to work in close proximity with human collaborators. A demonstration of our setup in which the human-robot team performs an assembly task in a shared but unknown environment, using algorithms for manipulation [5] and effective information sharing [10], is available at http://tiny.cc/hrc-ue1. We posit that despite differences in the knowledge representation used by humans and robots, by representing and learning latent human states a robot can improve coordination in human-robot teams.

References

[1] Stefano V. Albrecht and Peter Stone. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence*, 258:66 – 95, 2018.

[2] Maya Cakmak and Andrea L Thomaz. Designing robot learners that ask good questions. In *HRI*, 2012.

[3] Sonia Chernova and Andrea L Thomaz. Robot learning from human teachers. *Synthesis Lectures on AI and ML*, 8(3):1–121, 2014.

[4] Samuel J Gershman, Kenneth A Norman, and Yael Niv. Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Sciences*, 5:43–50, 2015.

[5] Charlie Guan, Will Vega-Brown, and Nicholas Roy. Efficient planning for near-optimal compliant manipulation leveraging environmental contact. In *ICRA*, 2018.

[6] Guy Hoffman and Cynthia Breazeal. Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team. In *HRI*, 2007.

[7] Stefania Pellegrinelli, Henny Admoni, Shervin Javdani, and Siddhartha Srinivasa. Human-robot shared workspace collaboration via hindsight optimization. In *IROS*, 2016.

[8] Dorsa Sadigh, Shankar Sastry, Sanjit A Seshia, and Anca D Dragan. Planning for Autonomous Cars that Leverage Effects on Human Actions. In *R:SS*, 2016.

[9] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

[10] Vaibhav V. Unhelkar and Julie A. Shah. ConTaCT: deciding to communicate during time-critical collaborative tasks in unknown, deterministic domains. In *AAAI*, 2016.

[11] Vaibhav V. Unhelkar*, Przemyslaw Lasota*, Quirin Tyroller, Rares-Darius Buhai, Marceau Laurie, Barbara Deml, and Julie A. Shah. Human-Aware Robotic Assistant for Collaborative Assembly: Integrating Human Motion Prediction with Planning in Time. *IEEE Robotics and Automation Letters (RA-L)*, 2018.

[12] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. Maximum Entropy Inverse Reinforcement Learning. In *AAAI*, 2008.