Detecting Hypothesis Space Misspecification in Robot Learning from Human Input

Andreea Bobu
University of California, Berkeley
abobu@berkeley.edu

ABSTRACT

Learning from human input has enabled autonomous agents to perform increasingly more complex tasks that are otherwise difficult to carry out automatically. To this end, recent works have studied how robots can incorporate such input – like demonstrations or corrections – into objective functions describing the desired behaviors. While these methods have shown progress in a variety of settings, from semi-autonomous driving, to household robotics, to automated airplane control, they all suffer from the same crucial drawback: *they implicitly assume that the person's intentions can always be captured by the robot's hypothesis space.* We call attention to the fact that this assumption is often unrealistic, as no model can completely account for every single possible situation ahead of time. When the robot's hypothesis space is *misspecified*, human input can be unhelpful – or even detrimental – to the way the robot is performing its tasks. Our work tackles this issue by proposing that the robot should first explicitly reason about how well its hypothesis space can explain human inputs, then use that *situational confidence* to inform how it should incorporate them.

KEYWORDS

Bayesian inference, inverse reinforcement learning

ACM Reference Format:
Andreea Bobu. 2020. Detecting Hypothesis Space Misspecification in Robot Learning from Human Input. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20 Companion), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3371382.3377436

1 INTRODUCTION

Imagine the household robotics scenario in Figure 1, where a human tries to get the robot to stay close to the table. If the robot's hypothesis space contains this preference, progress in learning from humans [1, 6–9] allows the robot to interpret the person's input and learn the correct hypothesis. However, if the robot's model does not capture distances from the table, the system can misinterpret human guidance, perform unexpected or undesired behavior, and degrade in overall performance. In these cases, we argue that the robot should *understand when it cannot understand* the input, instead of blindly learning unintended objectives from any interaction.

HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-7057-8/20/03.
https://doi.org/10.1145/3371382.3377436

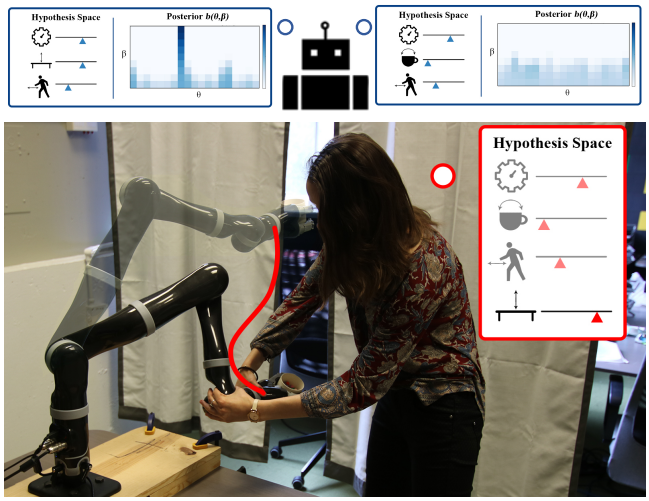

Figure 1: A person interacts with the robot, wanting it to be close to the table. (Top) When the robot's hypothesis space accounts for the table (left), its confidence β about what the input means is high for the correct hypothesis θ. When the robot's space is misspecified (right), β is low for all θs.

We tackled this issue in [4, 5] by introducing *situational confidence*, whereby the robot can quantify how much to trust its hypothesis space. Low confidence signals the robot should be cautious and request more guidance before proceeding, whereas high confidence suggests the robot can trust its model and learn assuredly.

2 SITUATIONAL CONFIDENCE FORMALISM

Notation. In our setup, a robot R assists a human H in the execution of some task. Both R and H can affect the evolution of the state $x \in \mathbb{R}^n$ over time via dynamics $x^{t+1} = f(x^t, u_R^t, u_H^t)$, where u_H and u_R are the control inputs. The result is a state trajectory $\mathbf{x} = [x^0, x^1, \ldots, x^T]$ given by a start state x^0 and the robot and human inputs $\mathbf{u}_R = [u_R^0, u_R^1, \ldots, u_R^T]$ and $\mathbf{u}_H = [u_H^0, u_H^1, \ldots, u_H^T]$.

The human has a preference ordering among trajectories given by a parametrized cost $C_\theta(\mathbf{x}, \mathbf{u}_R, \mathbf{u}_H)$, which is typically a function of features ϕ. R does not know C_θ, but it can use observations of the human input to draw inferences on θ. To do so, the robot needs an observation model describing how the human chooses her inputs. Following [3, 10], we model the human as a noisily-optimal agent exponentially likelier to choose actions with low cost:

$$P(\mathbf{u}_H \mid x^0, \mathbf{u}_R; \theta, \beta) = \frac{e^{-\beta C_\theta(\mathbf{x}, \mathbf{u}_R, \mathbf{u}_H)}}{\int_{\bar{\mathbf{u}}_H} e^{-\beta C_\theta(\mathbf{x}, \mathbf{u}_R, \bar{\mathbf{u}}_H)} d\bar{\mathbf{u}}_H}, \quad (1)$$

where $\beta \in [0, \infty)$ determines the degree to which the robot expects to observe human actions that are consistent with its cost model. Intuitively, $\beta \to 0$ models a randomly-acting human, while $\beta \to \infty$ models a perfectly optimal human.

Situational Confidence Estimation. Our goal is to detect when the robot's objective space cannot explain the human input. Different from regular cost learning, rather than only interpreting human input as evidence about *which* hypothesis is correct, we additionally focus on considering whether *any* hypothesis is correct. To tackle detecting hypothesis space misspecification, our insight is that we can reinterpret β as a *situational confidence* in the extent to which any hypothesis θ can explain the person's input. As such, when the space is correctly specified, human actions appear close to optimal, thus β is large; however, when the space is misspecified, human actions appear more random, corresponding to a low confidence β.

The robot can, thus, explicitly reason over its reliability of its human model in light of new evidence by maintaining a belief $b(\theta, \beta)$. For each new $\mathbf{u}_H$ given $x^0, \mathbf{u}_R$, this belief is updated as:

$$b'(\theta, \beta) = \frac{P(\mathbf{u}_H \mid x^0, \mathbf{u}_R; \theta, \beta) b(\theta, \beta)}{\int_{\bar{\theta}, \bar{\beta}} P(\mathbf{u}_H \mid x^0, \mathbf{u}_R; \bar{\theta}, \bar{\beta}) b(\bar{\theta}, \bar{\beta}) d\bar{\theta} d\bar{\beta}} \quad , \tag{2}$$

where $b'(\theta, \beta) = P(\theta, \beta \mid x^0, \mathbf{u}_R, \mathbf{u}_H)$.

Using β for robot learning. Given an estimate of this model confidence, there are many ways the robot could proceed in, depending on the context of its task. For example, in collaborative settings where a misunderstanding of the task's objective might be critical, the robot could stop and ask for clarification before proceeding. In other settings, such as when carrying out a known task but accommodating human preferences where possible, the robot could simply dismiss human inputs that result in low β values. Alternatively, the robot can plan to minimize the expected cost for the human given its current belief, by marginalizing over β:

$$\min_{\mathbf{u}_R} {}_{\theta \sim b} \left[C_\theta(\mathbf{x}, \mathbf{u}_R, \mathbf{u}_H) \right] \quad . \tag{3}$$

In our experiments, we focused on this latter situation given in (3), where the robot essentially learns in proportion to how confident it is in its ability to explain given human inputs.

3 EXPERIMENTAL RESULTS

We demonstrated our method's efficacy in detecting hypothesis space misspecification by running experiments and a user study on a 7 degree-of-freedom robotic manipulator learning from real human demonstrations and physical corrections.

3.1 Demonstrations

In [4], we collected 12 human demonstrations of household motion planning tasks and performed our situational confidence inference algorithm offline. As in Figure 1, we asked the participants to provide demonstrations with respect to a feature of interest, which the robot might (well-explained) or might not (poorly-explained) have in its hypothesis space. In situations where the input was well-explained (Figure 1, top-left), our Bayesian inference method would result in high β for the correct hypothesis θ. When taken separately, each demonstration would vary in the inferred situational confidence, ranging from high peaks as in the figure, to lower peaks for noisier inputs. However, when putting all user demonstrations together, the robot's inferred posterior would converge to a highly-confident and correct hypothesis. When the input was poorly-explained (Figure 1, top-right), our algorithm would always infer low βs for all hypotheses, regardless of the number of inputs.

3.2 Corrections

When learning from corrections, humans can intervene during the robot's task execution. As such, running the intractable inference in (2) is impractical for real-time use. As we detail in [5], we derive an online approximation of (2) to alleviate these computational challenges, which helps us separate β estimation from the θ update.

We ran an IRB-approved user study with 12 participants, where they were asked to physically correct the robot during its incorrect task execution. In two of the tasks, the humans were providing corrections that the robot's hypothesis space could capture. In the other two tasks, the robot's space was misspecified, which resulted in the participants having to repeatedly attempt to correct the robot. We compared our estimated confidence method to the state-of-the-art update that does not consider β [2]. The results of the study supported our hypotheses: for tasks where the robot's model was well-specified, our method was not inferior to the state-of-the-art and the participants felt like the methods performed the same; when the model was misspecified, our method reduced unintended learning significantly when compared to the baseline and our participants felt this difference.

4 FUTURE WORK

Our method introduces the situational confidence parameter β as a natural way to measure how much the robot should trust its explanations of the given inputs and learn from them. While this work is a good step in the direction of tackling misspecification, it has both limitations and possibilities for future work. The greatest drawback is that, in some cases, the hypothesis space will be misspecified but the robot will be able to explain the input nonetheless, thus confusing misspecification for noise. This is a fundamental problem with more expressive models in general, where there might always be some hypothesis that explains the input. While a single data point might not be enough, having more and diverse inputs lowers the chance of a single wrong hypothesis being able to explain well everything. An interesting research direction is analyzing how much data is enough, as well as how to modify the algorithm to discern misspecification from noise.

Another compelling next step is considering multiple hypothesis spaces, some more expressive than others, and switching between them whenever the situational confidence is very low for all θs. It would be valuable to analyze when some models, although more data- and computation-hungry, can perform better, and how often they run into the fundamental issue detailed above. Additionally, an interesting problem to explore is that of feature elicitation: how could the robot expand its hypothesis space online by actively querying the human? Indeed, automatically discovering human preferences and goals could have implications for reducing misspecification. Lastly, we are also interested in extending our work to sequential time-dependent inputs, since people often times can change their minds about which objectives they are considering.

REFERENCES

[1] Pieter Abbeel and Andrew Y Ng. 2004. Apprenticeship learning via inverse reinforcement learning. In *Machine Learning (ICML), International Conference on*. ACM.

[2] Andrea Bajcsy, Dylan P. Losey, Marcia Kilchenman O'Malley, and Anca D. Dragan. 2017. Learning Robot Objectives from Physical Human Interaction. In *CoRL*.

[3] Chris L Baker, Joshua B Tenenbaum, and Rebecca R Saxe. 2007. Goal inference as inverse planning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 29.

[4] Andreea Bobu, Andrea Bajcsy, Jaime F. Fisac, Sampada Deglurkar, and Anca D. Dragan. 2019. Quantifying Hypothesis Space Misspecification in Learning from Human-Robot Demonstrations and Physical Corrections. (2019). To appear in Transactions on Robotics.

[5] Andreea Bobu, Andrea Bajcsy, Jaime F. Fisac, and Anca D. Dragan. 2018. Learning under Misspecified Objective Spaces. In *2nd Annual Conference on Robot Learning, CoRL 2018, Zürich, Switzerland, 29-31 October 2018, Proceedings*. 796–805.

[6] Paul Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. (06 2017).

[7] Justin Fu, Avi Singh, Dibya Ghosh, Larry Yang, and Sergey Levine. [n.d.]. Variational Inverse Control with Events: A General Framework for Data-Driven Reward Definition. *arXiv preprint* arXiv:1805.11686 ([n. d.]).

[8] Ashesh Jain, Shikhar Sharma, Thorsten Joachims, and Ashutosh Saxena. 2015. Learning preferences for manipulation tasks from online coactive feedback. *The International Journal of Robotics Research* 34, 10 (2015), 1296–1313.

[9] Shervin Javdani, Siddhartha S Srinivasa, and J Andrew Bagnell. 2015. Shared autonomy via hindsight optimization. *arXiv preprint arXiv:1503.07619* (2015).

[10] John Von Neumann and Oskar Morgenstern. 1945. *Theory of games and economic behavior*. Princeton University Press Princeton, NJ.