Learning from Humans for Adaptive Interaction

Erdem Bıyık
Electrical Engineering, Stanford University
ebiyik@stanford.edu

***Abstract*—Robots that will cooperate (or even compete) with humans should understand their goals and preferences. Humans leak and provide a lot of data, *e.g.*, they take actions to achieve their goals, they make choices between multiple options, they use language or gestures to convey information. And we, as humans, are usually very good at using all these available information: we can easily understand what another person is trying to do just by watching them for a while. The goal of my research is to equip robots with the capability of using multiple modes of information sources. For this, I propose using a Bayesian learning approach, and show how it is useful in a variety of applications ranging from exoskeleton gait optimization to traffic routing.**

***Index Terms*—robot learning, human-in-the-loop learning**

I. Introduction

While robots and agents with artificial intelligence (AI) are increasingly becoming part of our lives, most of their current interactions with the humans is one-way, *e.g.*, a driver commands a vehicle to park autonomously, or the vehicle warns the driver about weather conditions. However, their successful integration into society will require them to intelligently *adapt to* and *influence* the humans and other robots.

These two-way interactions, where agents need to learn, adapt to, and influence each other; appear in almost all real-life scenarios. Human teams good at collaborating are often the ones where each individual adapted themselves to the others, *e.g.*, sports teams train together rather than trying to improve individually. However, AI agents are not yet capable of this adaptation: their inability to model others led to problems in several occasions. For example, price-setting bots tried to sell a book for $23.7M on an online retail website, because they were competing and did not realize if they increase the price, the other bot will also do that [1]. Though this is an old example, we still see similar issues arise: autonomous cars fail to change lanes as they do not know the other drivers will slow down if they simply nudge them [2]. My approach to enable robots to achieve the two-way interactions is inspired by how humans interact: we efficiently infer our partners' goals to optimize our behavior. For example, we move to one side of the sidewalk when we see a cyclist is approaching. If there is a mismatch between the inferred goal and our own goal, we try to influence our partners, *e.g.*, if the cyclist moves to the same side, we stop for a second to imply we want to stay on this side and they should use the other side.

My approach to equipping the robots with these capabilities consists of two parts. First, robots model the behaviors and goals of the other agents by learning from different forms of information they leak or explicitly provide. Second, they interact with the others to achieve online adaptation by leveraging the learned behaviors and goals, *e.g.*, an autonomous vehicle will adapt to both its driver and the other vehicles to better optimize its route and driving style.

In this article, I mostly focus on the first part: how can robots learn from and model humans? For this, I discuss learning from various forms of human feedback in Section II, and two interesting applications in Section III. Moving forward, I will do research on how these learned models of other agents can enable robots to adapt to and influence them. My long-term research goal is to enable robots to utilize all forms of information available in the environment, and use these models to achieve adaptation in complex multi-agent systems that even involve non-stationary agents or the need for online learning and teaching, some of which I discuss in Section IV.

II. Learning from Human Feedback

I propose using a Bayesian approach to learn from humans, where we learn their policy or reward function by updating our belief about them after observing every data sample. For example, if this policy or reward is parameterized by ω, then:

$$b^i(\omega) = b^{i-1}(\omega) p(\text{data}_i \mid \omega) \ ,$$

where b^{i-1} and b^i are the prior and the posterior belief about ω, and data_i is the observed data, *e.g.*, a demonstration or a choice made by the human. This approach requires only a model for the likelihood term $p(\text{data}_i \mid \omega)$. While the maximizer (or the expectation) of the posterior provides a good policy or reward, this approach also enables modeling uncertainties. Besides, even non-parametric models, *e.g.*, Gaussian processes, can be trained in this way as we showed in [3].

In addition to ability to incorporate new data, this Bayesian approach enables modeling uncertainties and learning *actively*. The former is important for safety-critical applications and understanding how good the human is modeled. The latter enables us to model humans quickly, which is often crucial in robotics, since data collection is very costly. Below, I give some examples of what kind of data we can learn from.

Learning from Demonstrations. Robots may learn through human demonstrations [4, 5]. Standard imitation learning and inverse reinforcement learning solutions suffer from the fact that when human demonstrations are suboptimal (which is often the case in robotics due to high degrees of freedom [6, 7, 8, 9], and cognitive biases [10, 11]), the robot cannot realize it. However as we showed in [12], the Bayesian learning approach enables us to keep a belief for the reward function and better tune it with other forms of data, *e.g.*, comparisons.

Learning from Comparisons. Demonstrations are only one information source, and are very sparse in robotics. Hence, we

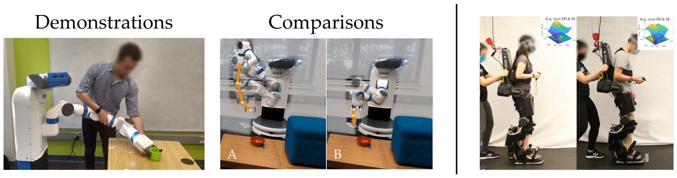

Fig. 1. In my research, I worked on **(left)** learning from various forms of human feedback, such as demonstrations and comparisons. **(right)** My work has enabled learning the gait preferences of the lower-body exoskeleton users.

focused on another source: comparisons [12], where a robot shows two different trajectories and the human selects which one they prefer. Comparisons are very reliable, as humans are much better at comparing options rather than finding optimal actions or quantifying success. With comparison feedback, we were able to have robots model humans' behavior by keeping a belief over their objectives [3, 12]. Besides, we developed algorithms that enable robots to learn *actively* by optimizing the comparison questions they ask to improve data-efficiency. To also increase time-efficiency, we developed batch-active methods for optimizing multiple questions at once while ensuring the diversity in questions to prevent redundancy [13]. We later released an open-source software library that unifies all these learning from comparisons methods [14].

Leveraging Richer Feedback. Our research showed the comparison questions can be enriched to improve both learning efficiency and expressiveness of the learned models. Specifically, we showed letting users indicate indifference between two options makes their responses more reliable, and leveraging this third option improves learning speed [15]. We further extended this work to enable AI agents learn from scale feedback where users tell how much they prefer one option over the other [16]. However, theoretical works proved pairwise comparisons are not enough to learn multimodal rewards which are often needed in real applications [17]. With the Bayesian learning approach, the comparisons need not be pairwise. We showed ranking queries where users rank multiple trajectories enable robots to learn multimodal rewards [18].

Importantly, all these improvements only require modeling a new likelihood term, $p(\text{data}_i \mid \omega)$. Hence, robots can use these different information sources together.

III. Applications

Before discussing some future works, I want to briefly mention two interesting applications of our learning algorithms.

First, they have had a direct impact on personalizing robots (see Fig. 1). We applied our learning from comparisons methods for lower-body exoskeletons [19], which aim to restore mobility to people with paralysis, a group with nearly 5.4 million people in the U.S. alone [20]. While these people cannot possibly provide demonstrations, they can give comparison feedback. Our active learning algorithms have also been crucial, as this is a very demanding process. With actively generated comparison queries, we have been able to understand their preferences and optimize comfort and safety.

Secondly, we noted people give choice feedback on ride-hailing applications, *e.g.*, Uber and Lyft, when they select between multiple commute options. Using these data, we learn their price-latency tradeoff, which enables us to come up with a joint pricing and routing strategy to reduce traffic congestion while still serving the same number of passengers [21, 22].

IV. Future Work

A robot with a good understanding of another agent's behavior can use this knowledge to predict what they are trying to do, assist them in their task, and even teach them how to perform the task better. To this end, I will discuss incorporating other forms of human feedback to improve the learned models, and how robots can best use these learned behaviors.

A. Other Forms of Human Feedback

I have so far focused on learning from comparisons, demonstrations, rankings, scale feedback, etc. However, humans may provide or leak information in many other ways, such as ordinal data [23], language [24], gaze [25] or gestures [26]. Future work may incorporate these into the Bayesian learning framework. Moreover, I think one of the ultimate goals for robot learning researchers should be developing foundation models [27] for robotics so that we can have robots that learn from all the available data in the environment despite the cost of getting explicit feedback from humans.

B. Incorporating Nonstationarity

In an earlier work, we showed hierarchical comparison questions, where users respond to a sequence queries whose trajectories follow each other, enable learning dynamically changing goals [28]. However, real users are more complex.

If a robot optimizes its behavior based on its predictions about others, they will start predicting what the robot will predict. This recursive reasoning, known as theory of mind, may go much deeper to the point where it is computationally intractable. Hence, robots should not always optimize their behavior with respect to the others. They should instead use game-theoretic techniques to find and reach the equilibrium that is best for them. Future work may focus on finding and reaching these equilibria in multi-agent environments.

Another nonstationarity is due to humans' latent states. As an example, how much a human trusts a robot varies over time based on the task performance, and is not directly observable. Future work may formulate the learning problem with a partially observable Markov decision process to predict these latent states while also learning the humans' objectives.

C. Teaching Humans

A robot with a good understanding of a human's behavior can use this knowledge to teach them. As an example, computers have been better than humans in chess for more than 20 years. Players have started to use these chess engines to create better openings, lessons, or chess puzzles [29]. I believe similar ideas may be useful for more complex and dynamical systems. For example, can we have a semi-autonomous vehicle that makes humans better drivers? Or can we replace the training wheels of a bike with a self-balancing system to teach how to bike to beginners?

References

[1] John D Sutter. "Amazon seller lists book at $23,698,655.93 – plus shipping". In: *CNN* (Apr. 2011). URL: http://www.cnn.com/2011/TECH/web/04/25/amazon.price.algorithm/index.html.

[2] Daily Mail. *Waymo's self driving minivans struggles to merge in left lane*. 2020. URL: https://www.dailymail.co.uk/video/sciencetech/video-1752896/Video-Waymos-self-driving-minivans-struggles-merge-left-lane.html.

[3] Erdem Biyik et al. "Active Preference-Based Gaussian Process Regression for Reward Learning". In: *Proceedings of Robotics: Science and Systems (RSS)*. July 2020. DOI: 10.15607/rss.2020.xvi.041.

[4] Deepak Ramachandran and Eyal Amir. "Bayesian Inverse Reinforcement Learning." In: *IJCAI*. Vol. 7. 2007, pp. 2586–2591.

[5] Brian D Ziebart et al. "Maximum entropy inverse reinforcement learning." In: *Aaai*. Vol. 8. Chicago, IL, USA. 2008, pp. 1433–1438.

[6] Baris Akgun et al. "Keyframe-based learning from demonstration". In: *International Journal of Social Robotics* 4.4 (2012), pp. 343–355.

[7] Anca D Dragan and Siddhartha S Srinivasa. *Formalizing assistive teleoperation*. MIT Press, July, 2012.

[8] Shervin Javdani, Siddhartha S Srinivasa, and J Andrew Bagnell. "Shared autonomy via hindsight optimization". In: *Robotics science and systems: online proceedings* 2015 (2015).

[9] Rebecca P Khurshid and Katherine J Kuchenbecker. "Data-driven motion mappings improve transparency in teleoperation". In: *Presence* 24.2 (2015), pp. 132–154.

[10] Minae Kwon et al. "When Humans Aren't Optimal: Robots that Collaborate with Risk-Aware Humans". In: *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Mar. 2020. DOI: 10.1145/3319502.3374832.

[11] Chandrayee Basu et al. "Do you want your autonomous car to drive like you?" In: *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE. 2017, pp. 417–425.

[12] Erdem Biyik et al. "Learning Reward Functions from Diverse Sources of Human Feedback: Optimally Integrating Demonstrations and Preferences". In: *The International Journal of Robotics Research (IJRR)* (2021). DOI: 10.1177/02783649211041652.

[13] Erdem Biyik and Dorsa Sadigh. "Batch Active Preference-Based Learning of Reward Functions". In: *Proceedings of the 2nd Conference on Robot Learning (CoRL)*. Vol. 87. Proceedings of Machine Learning Research. PMLR, 2018, pp. 519–528.

[14] Erdem Biyik, Aditi Talati, and Dorsa Sadigh. "APReL: A Library for Active Preference-based Reward Learning Algorithms". In: *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2022.

[15] Erdem Biyik et al. "Asking Easy Questions: A User-Friendly Approach to Active Reward Learning". In: *Proceedings of the 3rd Conference on Robot Learning (CoRL)*. 2019.

[16] Nils Wilde et al. "Learning Reward Functions from Scale Feedback". In: *Proceedings of the 5th Conference on Robot Learning (CoRL)*. 2021.

[17] Zhibing Zhao, Peter Piech, and Lirong Xia. "Learning mixtures of Plackett-Luce models". In: *International Conference on Machine Learning*. PMLR. 2016, pp. 2906–2914.

[18] Vivek Myers et al. "Learning Multimodal Rewards from Rankings". In: *Proceedings of the 5th Conference on Robot Learning (CoRL)*. 2021.

[19] Kejun Li et al. "ROIAL: Region of Interest Active Learning for Characterizing Exoskeleton Gait Preference Landscapes". In: *International Conference on Robotics and Automation (ICRA)*. May 2021. DOI: 10.1109/ICRA48506.2021.9560840.

[20] Brian S Armour et al. "Prevalence and causes of paralysis—United States, 2013". In: *American journal of public health* 106.10 (2016), pp. 1855–1857.

[21] Erdem Biyik et al. "The Green Choice: Learning and Influencing Human Decisions on Shared Roads". In: *Proceedings of the 58th IEEE Conference on Decision and Control (CDC)*. Dec. 2019. DOI: 10.1109/CDC40024.2019.9030169.

[22] Erdem Biyik et al. "Incentivizing Efficient Equilibria in Traffic Networks with Mixed Autonomy". In: *IEEE Transactions on Control of Network Systems (TCNS)* (2021). DOI: 10.1109/TCNS.2021.3084045.

[23] Wei Chu, Zoubin Ghahramani, and Christopher KI Williams. "Gaussian processes for ordinal regression." In: *Journal of machine learning research* 6.7 (2005).

[24] Dilip Arumugam et al. "Grounding natural language instructions to semantic goal representations for abstraction and generalization". In: *Autonomous Robots* 43.2 (2019), pp. 449–468.

[25] Javier L Castellanos, Maria F Gomez, and Kim D Adams. "Using machine learning based on eye gaze to predict targets: An exploratory study". In: *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE. 2017, pp. 1–7.

[26] Andrea Lockerd Thomaz, Guy Hoffman, and Cynthia Breazeal. "Real-time interactive reinforcement learning for robots". In: *AAAI 2005 workshop on human comprehensible machine learning*. 2005.

[27] Rishi Bommasani et al. "On the opportunities and risks of foundation models". In: *arXiv preprint arXiv:2108.07258* (2021).

[28] Chandrayee Basu et al. "Active Learning of Reward Dynamics from Hierarchical Queries". In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Nov. 2019. DOI: 10.1109/IROS40897.2019.8968522.

[29] Linmiao Xu and Nicolas Lazo. *Chess Puzzle Maker*. https://github.com/linrock/chess-puzzle-maker. 2020.