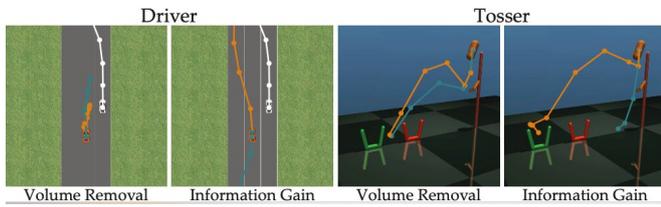


# A user-friendly approach for active reward learning in robots

24 October 2019, by Ingrid Fadelli



Credit: B?y?k et al.

In recent years, researchers have been trying to develop methods that enable robots to learn new skills. One option is for a robot to learn these new skills from humans, asking questions whenever it is unsure about how to behave, and learning from the human user's responses.

A research team at Stanford University recently developed a user-friendly approach to active reward learning that can be used to train robots by having human users answer their questions. This new approach, presented in [a paper prepublished on arXiv](#), trains robots to ask questions that will be easy for a human user to answer and that are not redundant or unnecessary.

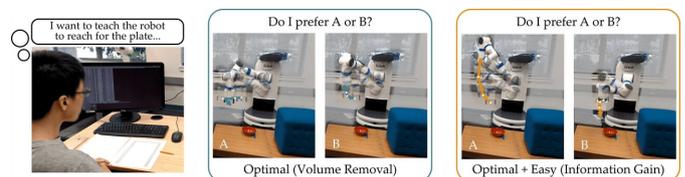
"Our group is interested in how robots can learn what humans want," the researchers told TechXplore via email. "One intuitive way to learn is by asking questions. For example, would you rather an autonomous car drive cautiously or aggressively? Should this autonomous car merge in front of or behind a human-driven car?"

The main assumption behind the recent study is that ideally, robots should ask informative questions that elicit as much information as possible from human users. In other words, a robot should be able to understand what a human needs or wants them to do by asking as few questions as possible.

In reality, however, most existing training approaches based on question answering do not consider how easy it will be for human users to answer specific questions formulated by the robot. This often results in users wasting their time answering loads of unnecessary questions or being unable to respond with certainty.

"We found that most state-of-the-art algorithms show the human alternatives that are (almost) indistinguishable, preventing the person from correctly answering the robot's questions," the researchers said. "Returning to our example, these approaches might ask: "Would you rather merge in front of the human-driven car at a speed of 29 mph, or a speed of 31 mph?" This can be informative for the robot to decide whether the human wants to go faster than 30 mph or not, but the options are so close that humans cannot reliably respond."

To overcome the limitations of existing active learning methods, the researchers developed an algorithm that can select more effective questions to ask human users. The algorithm identifies questions that most reduce the robot's uncertainty about a human user's preferences (i.e., that maximize information gain), while also considering how easy it will be for a human user to answer them.



Credit: B?y?k et al.

"Inspired by the shortcomings of prior works, when we developed this algorithm, we focused on

accounting for the human's ability to actually answer techniques."

the questions that the robot is asking," the researchers said. "This is based on the idea that only robots that account for the human's ability to answer can accurately and efficiently learn what humans want."

The researchers calculated information gain by measuring the decrease in entropy (i.e., a measure of uncertainty) over the human user's preferences as a function of the question asked by the robot. In other words, a question that maximizes information gain will most reduce the robot's uncertainty over what the human user's preferences are. This gives robots a formal objective that they can use to select questions that are most informative.

"One nice characteristic of information gain is that it inherently maximizes the robot's uncertainty (so that the robot learns a lot from the question) while also minimizing the human's uncertainty (so that the question is easy for the human to answer)," the researchers explained. "Generating the questions using information gain thus improves active learning, not only because the questions are maximally informative, but also because the human gives fewer erroneous responses."

The approach devised by the researchers greedily selects the question that maximizes information gain at every time step. Essentially, the robot maintains a belief (i.e., a probability distribution) over the preferences of the user it is interacting with and samples from both this belief and the space of possible questions.

Ultimately, the robot chooses the question that provides the most information gain across the current distribution of possible human preferences. Subsequently, it updates its beliefs about what the user wants based on the answer it receives. This process is continuously repeated, allowing the robot to gradually improve its performance by learning about the user's preferences.

"We formulated a computationally tractable method that allows us to quickly discover human preferences on real robotic tasks, outperforming prior methods," the researchers said. "In our study, users preferred our method to other state-of-the-art

In their study, the Stanford-based team showed that training a robot to ask questions that maximize information gain has the same computational complexity as state-of-the-art methods. In other words, it is not any harder for the robot to find these informative questions, compared to those generated by other approaches.

"We also point out that our approach has several desirable mathematical properties, such as submodularity, which enables us to take the extensions and theoretical bounds that were developed for prior approaches and also use them with our method," the researchers said. "For example, we can use prior works to find several informative questions at once, instead of searching for one question at a time."

The team evaluated their active reward-learning approach in a series of simulations and found that it allows robots to grasp human preferences faster and more accurately than other state-of-the-art methods. This was also found to be true in situations in which humans can correctly answer difficult questions or when their answer is "I don't know."

The researchers also carried out a user study in which they asked human participants to answer questions generated by their method and others generated using other state-of-the-art approaches. The feedback they collected suggests that people find questions generated by their approach far easier to answer. In addition, users often felt that robots using the new method had acquired a more accurate representation of their preferences than they did with previously proposed approaches.

"Considering all of our contributions together, we took a step toward enabling robots to determine human preferences," the researchers said. "We showed that the true objective that we originally wanted the [robot](#) to maximize—asking questions to gain as much information as possible—can actually be solved with the same computational complexity as existing methods."

In the future, the active reward-learning technique

developed by this team of researchers could help to train robots more effectively, making them more attuned to user preferences. In addition, it could be used to teach robots to ask questions that humans can easily understand and answer. In their future studies, the researchers would also like to investigate methods for training robots to give useful explanations for their actions.

"We are excited about robots that not only ask good questions, but can also explain why they are asking those questions," the researchers said. "We imagine a scenario where a self-driving car visualizes two different merging options for the human, and then clarifies that it is asking about these options because it is rush hour, and it wants to determine whether it should behave more or less aggressively."

**More information:** Asking easy questions: A user-friendly approach to active reward learning.

arXiv:1910.04365 [cs.RO].

[arxiv.org/abs/1910.04365](https://arxiv.org/abs/1910.04365)

© 2019 Science X Network

APA citation: A user-friendly approach for active reward learning in robots (2019, October 24) retrieved 5 December 2020 from <https://techxplore.com/news/2019-10-user-friendly-approach-reward-robots.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*