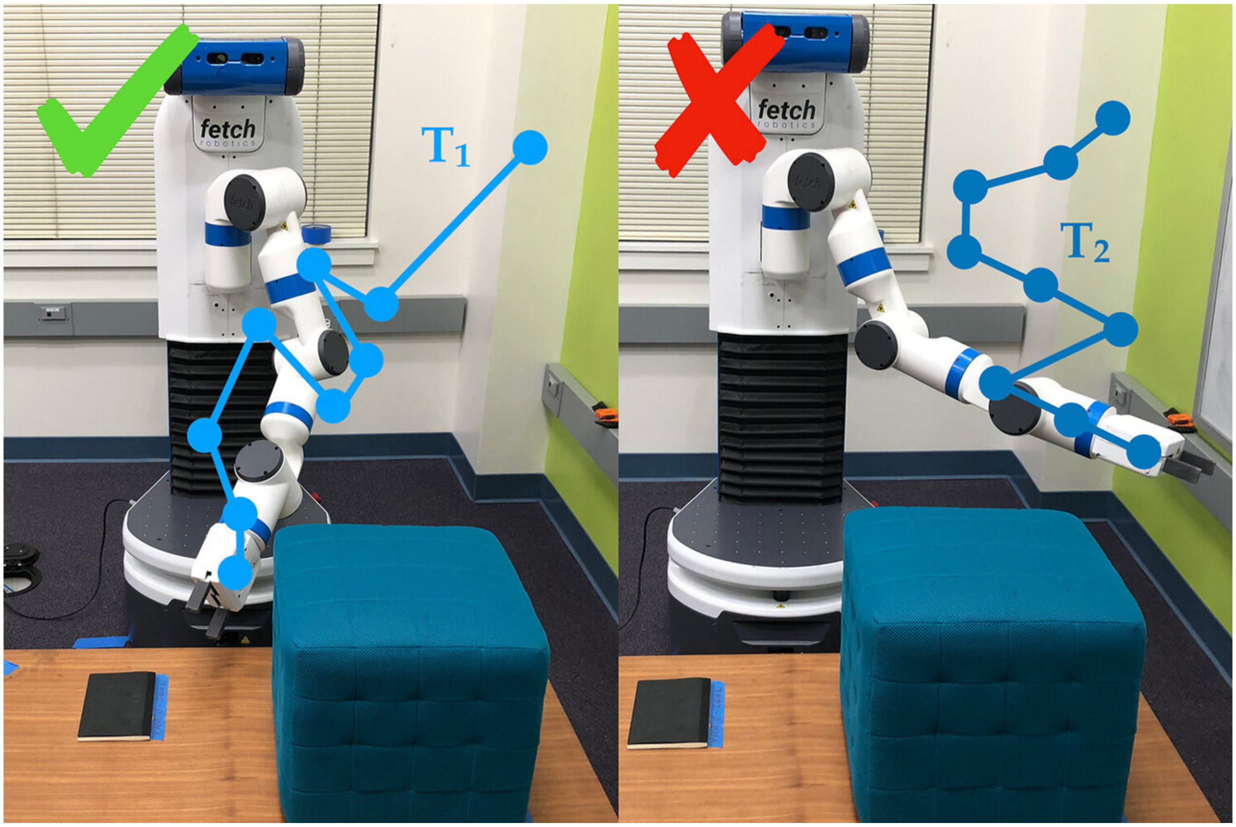
TechXplore

Teaching robots what humans want

June 24 2019, by Taylor Kubota



An example of how the robot arm uses survey questions to determine the preferences of the person using it. In this case, the person prefers trajectory #1 (T1) over trajectory #2. Credit: Andy Palan and Gleb Shevchuk

Told to optimize for speed while racing down a track in a computer game, a car pushes the pedal to the metal … and proceeds to spin in a tight little circle. Nothing in the instructions told the car to drive straight, and so it improvised.

This example—funny in a computer game but not so much in life—is among those that motivated Stanford University researchers to build a better way to set goals for autonomous systems.

Dorsa Sadigh, assistant professor of computer science and of electrical engineering, and her lab have combined two different ways of setting goals for robots into a single process, which performed better than either of its parts alone in both simulations and real-world experiments. The researchers presented the work June 24 at the *Robotics: Science and Systems* conference.

"In the future, I fully expect there to be more autonomous systems in the world and they are going to need some concept of what is good and what is bad," said Andy Palan, graduate student in computer science and co-lead author of the paper. "It's crucial, if we want to deploy these autonomous systems in the future, that we get that right."

The team's new system for providing instruction to robots—known as reward functions—combines demonstrations, in which humans show the robot what to do, and user preference surveys, in which people answer questions about how they want the robot to behave.

"Demonstrations are informative but they can be noisy. On the other hand, preferences provide, at most, one bit of information, but are way more accurate," said Sadigh. "Our goal is to get the best of both worlds, and combine data coming from both of these sources more intelligently to better learn about humans' preferred reward function."

Demonstrations and surveys

In previous work, Sadigh had focused on preference surveys alone.

These ask people to compare scenarios, such as two trajectories for an autonomous car. This method is efficient, but could take as much as three minutes to generate the next question, which is still slow for creating instructions for complex systems like a car.

To speed that up, the group later developed a way of producing multiple questions at once, which could be answered in quick succession by one person or distributed among several people. This update sped the process 15 to 50 times compared to producing questions one-by-one.

The new combination system begins with a person demonstrating a behavior to the robot. That can give autonomous robots a lot of information, but the robot often struggles to determine what parts of the demonstration are important. People also don't always want a robot to behave just like the human that trained it.

"We can't always give demonstrations, and even when we can, we often can't rely on the information people give," said Erdem Biyik, a graduate student in electrical engineering who led the work developing the multiple-question surveys. "For example, previous studies have shown people want autonomous cars to drive less aggressively than they do themselves."

That's where the surveys come in, giving the robot a way of asking, for example, whether the user prefers it move its arm low to the ground or up toward the ceiling. For this study, the group used the slower single question method, but they plan to integrate multiple-question surveys in later work.

In tests, the team found that combining demonstrations and surveys was faster than just specifying preferences and, when compared with demonstrations alone, about 80 percent of people preferred how the robot behaved when trained with the combined system.

"This is a step in better understanding what people want or expect from a robot," said Sadigh. "Our work is making it easier and more efficient for humans to interact and teach robots, and I am excited about taking this work further, particularly in studying how robots and humans might learn from each other."

Better, faster, smarter

People who used the combined method reported difficulty understanding what the system was getting at with some of its questions, which sometimes asked them to select between two scenarios that seemed the same or seemed irrelevant to the task—a common problem in preference-based learning. The researchers are hoping to address this shortcoming with easier surveys that also work more quickly.

"Looking to the future, it's not 100 percent obvious to me what the right way to make reward functions is, but realistically you're going to have some sort of combination that can address complex situations with human input," said Palan. "Being able to design reward functions for autonomous systems is a big, important problem that hasn't received quite the attention in academia as it deserves."

The team is also interested in a variation on their system, which would allow people to simultaneously create reward functions for different scenarios. For example, a person may want their car to drive more conservatively in slow traffic and more aggressively when traffic is light.

When demos fail

Sometimes demonstrations alone fail to convey the point of a task. For example, one demonstration in this study had people teach the robot arm to move until it pointed at a specific spot on the ground, and to do that while avoiding an obstacle and without moving above a certain height.

After a human ran the robot through its paces for 30 minutes, the robot tried to perform the task autonomously. It simply pointed straight up. It was so focused on learning not to hit the obstacle, it completely missed the actual goal of the task—pointing to the spot—and the preference for staying low.

Hand coding and reward hacking

Another way to teach a robot is to write code that acts as instructions. The challenge is explaining exactly what you want a robot to do, especially if the task is complex. A common problem is known as "reward hacking," where the robot figures out an easier way to reach the specified goals—such as the car spinning in circles in order to achieve the goal of going fast.

Biyik experienced reward hacking when he was programming a robot arm to grasp a cylinder and hold it in the air.

"I told it the hand must be closed, the object has to have height higher than X and the hand should be at the same height," described Biyik. "The robot rolled the cylinder object to the edge of the table, hit it upward and then made a fist next to it in the air."

More information: Malayandi Palan et al. Learning Reward Functions by Integrating Human Demonstrations and Preferences. *Robotics: Science and Systems*, 2019. www.roboticsproceedings.org/rss15/p23.pdf

Provided by Stanford University

Citation: Teaching robots what humans want (2019, June 24) retrieved 25 April 2024 from https://techxplore.com/news/2019-06-robots-humans.html