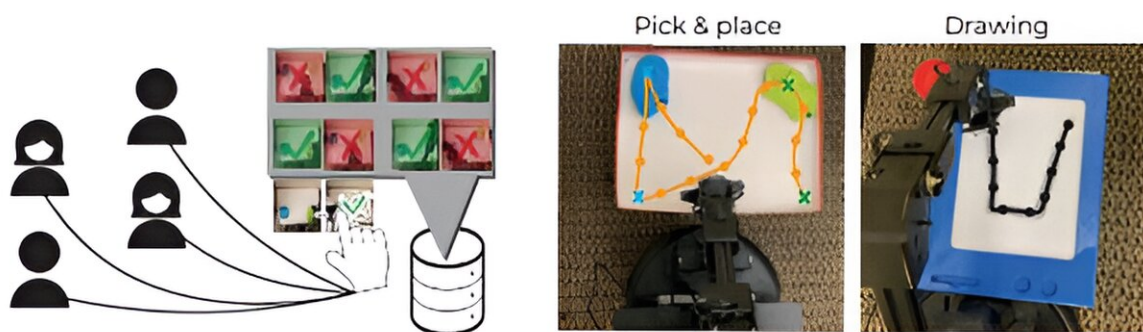# New method uses crowdsourced feedback to train robots

November 27 2023, by Adam Zewe

HuGE leverages noisy and asynchronous feedback from multiple non-expert humans to train robot control policies directly on the real world. Credit: *arXiv* (2023). DOI: 10.48550/arxiv.2307.11049

To teach an AI agent a new task, like how to open a kitchen cabinet, researchers often use reinforcement learning—a trial-and-error process where the agent is rewarded for taking actions that get it closer to the goal.

In many instances, a human expert must carefully design a reward function, which is an incentive mechanism that gives the agent motivation to explore. The human expert must iteratively update that reward function as the agent explores and tries different actions. This can be time-consuming, inefficient, and difficult to scale up, especially

when the task is complex and involves many steps.

Researchers from MIT, Harvard University, and the University of Washington have developed a new reinforcement learning approach that doesn't rely on an expertly designed reward function. Instead, it leverages crowdsourced feedback, gathered from many non-expert users, to guide the agent as it learns to reach its goal. The work has been published on the pre-print server *arXiv*.

While some other methods also attempt to utilize non-expert feedback, this new approach enables the AI agent to learn more quickly, despite the fact that data crowdsourced from users are often full of errors. These noisy data might cause other methods to fail.

In addition, this new approach allows feedback to be gathered asynchronously, so non-expert users around the world can contribute to teaching the agent.

"One of the most time-consuming and challenging parts in designing a robotic agent today is engineering the reward function. Today reward functions are designed by expert researchers—a paradigm that is not scalable if we want to teach our robots many different tasks. Our work proposes a way to scale robot learning by crowdsourcing the design of reward function and by making it possible for non-experts to provide useful feedback," says Pulkit Agrawal, an assistant professor in the MIT Department of Electrical Engineering and Computer Science (EECS) who leads the Improbable AI Lab in the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL).

In the future, this method could help a robot learn to perform specific tasks in a user's home quickly, without the owner needing to show the robot physical examples of each task. The robot could explore on its own, with crowdsourced non-expert feedback guiding its exploration.

"In our method, the reward function guides the agent to what it should explore, instead of telling it exactly what it should do to complete the task. So, even if the human supervision is somewhat inaccurate and noisy, the agent is still able to explore, which helps it learn much better," explains lead author Marcel Torne, a research assistant in the Improbable AI Lab.

Torne is joined on the paper by his MIT advisor, Agrawal; senior author Abhishek Gupta, assistant professor at the University of Washington; as well as others at the University of Washington and MIT. The research will be presented at the Conference on Neural Information Processing Systems next month.

Noisy feedback

One way to gather user feedback for reinforcement learning is to show a user two photos of states achieved by the agent, and then ask users to state which is closer to a goal. For instance, perhaps a robot's goal is to open a kitchen cabinet. One image might show that the robot opened the cabinet, while the second might show that it opened the microwave. A user would pick the photo of the "better" state.

Some previous approaches try to use this crowdsourced, binary feedback to optimize a reward function that the agent would use to learn the task. However, because non-experts are likely to make mistakes, the reward function can become very noisy, so the agent might get stuck and never reach its goal.

"Basically, the agent would take the reward function too seriously. It would try to match the reward function perfectly. So, instead of directly optimizing over the reward function, we just use it to tell the robot which areas it should be exploring," Torne says.

He and his collaborators decoupled the process into two separate parts, each directed by its own algorithm. They call their new reinforcement learning method HuGE (Human Guided Exploration).

On one side, a goal selector algorithm is continuously updated with crowdsourced human feedback. The feedback is not used as a reward function, but rather to guide the agent's exploration. In a sense, the non-expert users drop breadcrumbs that incrementally lead the agent toward its goal.

On the other side, the agent explores on its own, in a self-supervised manner guided by the goal selector. It collects images or videos of actions that it tries, which are then sent to humans and used to update the goal selector.

This narrows down the area for the agent to explore, leading it to more promising areas that are closer to its goal. But if there is no feedback, or if feedback takes a while to arrive, the agent will keep learning on its own, albeit in a slower manner. This enables feedback to be gathered infrequently and asynchronously.

"The exploration loop can keep going autonomously, because it is just going to explore and learn new things. And then when you get some better signal, it is going to explore in more concrete ways. You can just keep them turning at their own pace," adds Torne.

And because the feedback is just gently guiding the agent's behavior, it will eventually learn to complete the task even if users provide incorrect answers.

Faster learning

The researchers tested this method on a number of simulated and real-

world tasks. In simulation, they used HuGE to effectively learn tasks with long sequences of actions, such as stacking blocks in a particular order or navigating a large maze.

In real-world tests, they utilized HuGE to train robotic arms to draw the letter "U" and pick and place objects. For these tests, they crowdsourced data from 109 non-expert users in 13 different countries spanning three continents.

In real-world and simulated experiments, HuGE helped agents learn to achieve the goal faster than other methods.

The researchers also found that data crowdsourced from non-experts yielded better performance than synthetic data, which were produced and labeled by the researchers. For non-expert users, labeling 30 images or videos took fewer than two minutes.

"This makes it very promising in terms of being able to scale up this method," Torne adds.

In a related paper, which the researchers presented at the recent Conference on Robot Learning, they enhanced HuGE so an AI agent can learn to perform the task, and then autonomously reset the environment to continue learning. For instance, if the agent learns to open a cabinet, the method also guides the agent to close the cabinet.

"Now we can have it learn completely autonomously without needing human resets," he says.

The researchers also emphasize that, in this and other learning approaches, it is critical to ensure that AI agents are aligned with human values.

In the future, they want to continue refining HuGE so the agent can learn from other forms of communication, such as natural language and physical interactions with the robot. They are also interested in applying this method to teach multiple agents at once.

More information: Marcel Torne et al, Breadcrumbs to the Goal: Goal-Conditioned Exploration from Human-in-the-Loop Feedback, *arXiv* (2023). DOI: 10.48550/arxiv.2307.11049