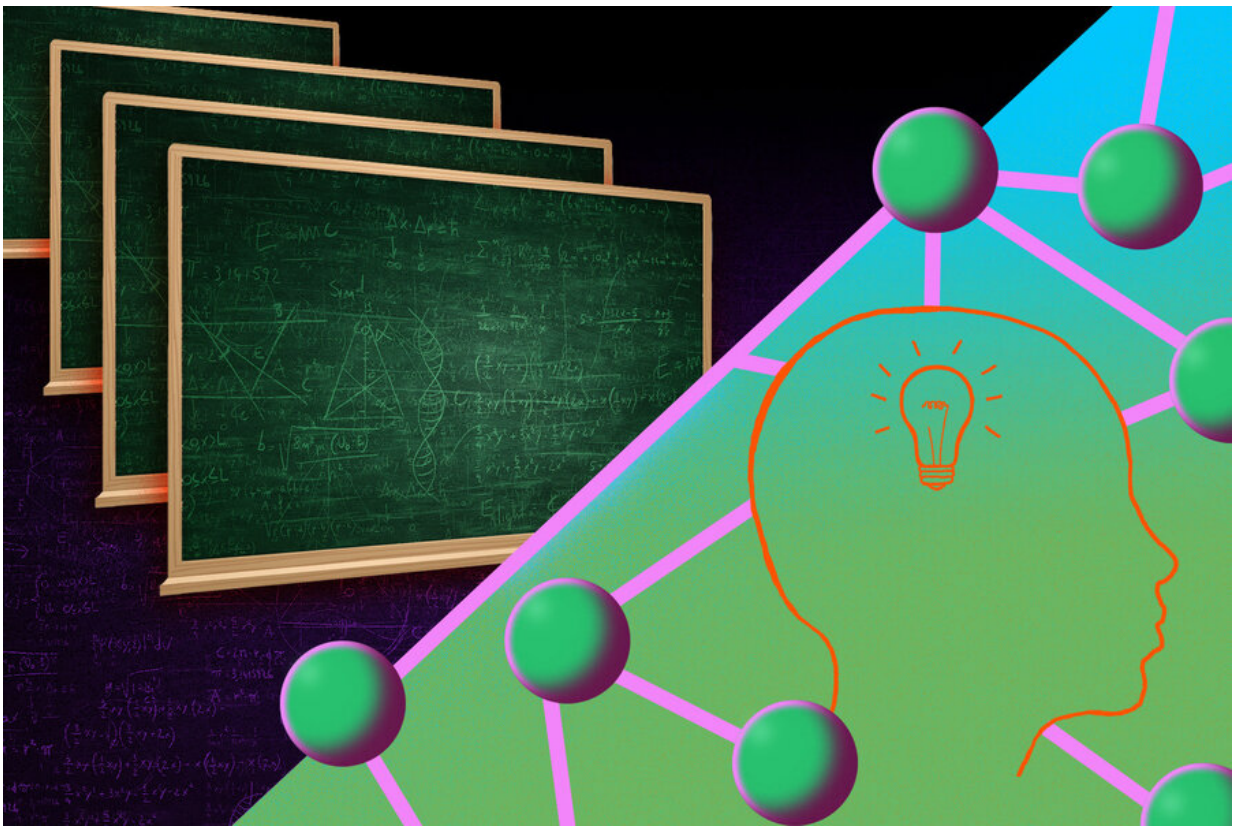
A more effective way to train machines for uncertain, real-world situations

May 31 2023, by Adam Zewe



Researchers from MIT and elsewhere developed an algorithm that automatically and dynamically determines whether a machine learning to complete a task should try to mimic its teacher or explore on its own through trial-and-error. This algorithm enabled simulated student machines to learn tasks faster and more effectively than other techniques. Credit: Jose-Luis Olivares/MIT

Someone learning to play tennis might hire a teacher to help them learn faster. Because this teacher is (hopefully) a great tennis player, there are times when trying to exactly mimic the teacher won't help the student learn. Perhaps the teacher leaps high into the air to deftly return a volley. The student, unable to copy that, might instead try a few other moves on her own until she has mastered the skills she needs to return volleys.

Computer scientists can also use "teacher" systems to train another machine to complete a task. But just like with human learning, the student machine faces a dilemma of knowing when to follow the teacher and when to explore on its own. To this end, researchers from MIT and Technion, the Israel Institute of Technology, have developed an algorithm that automatically and independently determines when the student should mimic the teacher (known as imitation learning) and when it should instead learn through trial and error (known as reinforcement learning).

Their dynamic approach allows the student to diverge from copying the teacher when the teacher is either too good or not good enough, but then return to following the teacher at a later point in the training process if doing so would achieve better results and faster learning.

When the researchers tested this approach in simulations, they found that their combination of trial-and-error learning and imitation learning enabled students to learn tasks more effectively than methods that used only one type of learning.

This method could help researchers improve the training process for machines that will be deployed in uncertain real-world situations, like a robot being trained to navigate inside a building it has never seen before.

"This combination of learning by trial-and-error and following a teacher is very powerful. It gives our algorithm the ability to solve very difficult

tasks that cannot be solved by using either technique individually," says Idan Shenfeld an electrical engineering and computer science (EECS) graduate student and lead author of a paper on this technique.

Shenfeld wrote the paper with co-authors Zhang-Wei Hong, an EECS graduate student; Aviv Tamar; assistant professor of electrical engineering and computer science at Technion; and senior author Pulkit Agrawal, director of Improbable AI Lab and an assistant professor in the Computer Science and Artificial Intelligence Laboratory. The research will be presented at the International Conference on Machine Learning.

Striking a balance

Many existing methods that seek to strike a balance between imitation learning and reinforcement learning do so through brute force trial-and-error. Researchers pick a weighted combination of the two learning methods, run the entire training procedure, and then repeat the process until they find the optimal balance. This is inefficient and often so computationally expensive it isn't even feasible.

"We want algorithms that are principled, involve tuning of as few knobs as possible, and achieve high performance—these principles have driven our research," says Agrawal.

To achieve this, the team approached the problem differently than prior work. Their solution involves training two students: one with a weighted combination of reinforcement learning and imitation learning, and a second that can only use reinforcement learning to learn the same task.

The main idea is to automatically and dynamically adjust the weighting of the reinforcement and imitation learning objectives of the first student. Here is where the second student comes into play. The researchers' algorithm continually compares the two students. If the one

using the teacher is doing better, the algorithm puts more weight on imitation learning to train the student, but if the one using only trial and error is starting to get better results, it will focus more on learning from reinforcement learning.

By dynamically determining which method achieves better results, the algorithm is adaptive and can pick the best technique throughout the training process. Thanks to this innovation, it is able to more effectively teach students than other methods that aren't adaptive, Shenfeld says.

"One of the main challenges in developing this algorithm was that it took us some time to realize that we should not train the two students independently. It became clear that we needed to connect the agents to make them share information, and then find the right way to technically ground this intuition," Shenfeld says.

Solving tough problems

To test their approach, the researchers set up many simulated teacher-student training experiments, such as navigating through a maze of lava to reach the other corner of a grid. In this case, the teacher has a map of the entire grid while the student can only see a patch in front of it. Their algorithm achieved an almost perfect success rate across all testing environments, and was much faster than other methods.

To give their algorithm an even more difficult test, they set up a simulation involving a robotic hand with touch sensors but no vision, that must reorient a pen to the correct pose. The teacher had access to the actual orientation of the pen, while the student could only use touch sensors to determine the pen's orientation.

Their method outperformed others that used either only imitation learning or only reinforcement learning.

Reorienting objects is one among many manipulation tasks that a future home robot would need to perform, a vision that the Improbable AI lab is working toward, Agrawal adds.

Teacher-student learning has successfully been applied to train robots to perform complex object manipulation and locomotion in simulation and then transfer the learned skills into the real-world. In these methods, the teacher has privileged information accessible from the simulation that the student won't have when it is deployed in the real world. For example, the teacher will know the detailed map of a building that the student robot is being trained to navigate using only images captured by its camera.

"Current methods for student-teacher learning in robotics don't account for the inability of the student to mimic the teacher and thus are performance-limited. The new method paves a path for building superior robots," says Agrawal.

Apart from better robots, the researchers believe their algorithm has the potential to improve performance in diverse applications where imitation or reinforcement learning is being used. For example, large language models such as GPT-4 are very good at accomplishing a wide range of tasks, so perhaps one could use the large model as a teacher to train a smaller, student model to be even "better" at one particular task. Another exciting direction is to investigate the similarities and differences between machines and humans learning from their respective teachers. Such analysis might help improve the learning experience, the researchers say.

"What's interesting about [this method] compared to related methods is how robust it seems to various parameter choices, and the variety of domains it shows promising results in," says Abhishek Gupta, an assistant professor at the University of Washington, who was not

involved with this work. "While the current set of results are largely in simulation, I am very excited about the future possibilities of applying this work to problems involving memory and reasoning with different modalities such as tactile sensing."

"This work presents an interesting approach to reuse prior computational work in reinforcement learning. Particularly, their proposed method can leverage suboptimal teacher policies as a guide while avoiding careful hyperparameter schedules required by prior methods for balancing the objectives of mimicking the teacher versus optimizing the task reward," adds Rishabh Agarwal, a senior research scientist at Google Brain, who was also not involved in this research. "Hopefully, this work would make reincarnating reinforcement learning with learned policies less cumbersome."

More information: TGRL: An Algorithm for Teacher Guided Reinforcement Learning. openreview.net/pdf?id=kTqjkIvjj7

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: A more effective way to train machines for uncertain, real-world situations (2023, May 31) retrieved 5 May 2024 from https://techxplore.com/news/2023-05-effective-machines-uncertain-real-world-situations.html