

Using imitation and reinforcement learning to tackle long-horizon robotic tasks

November 13 2019, by Ingrid Fadelli



Credit: Gupta et al.

Reinforcement learning (RL) is a widely used machine-learning

technique that entails training AI agents or robots using a system of reward and punishment. So far, researchers in the field of robotics have primarily applied RL techniques in tasks that are completed over relatively short periods of time, such as moving forward or grasping objects.

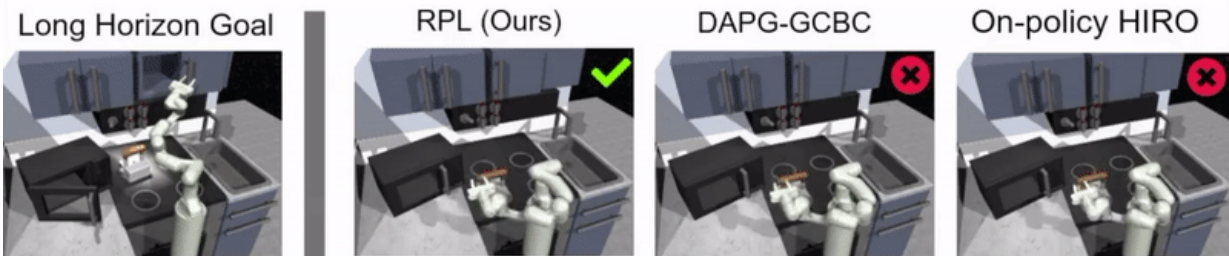
A team of researchers at Google and Berkeley AI Research has recently developed a new approach that combines RL with learning by imitation, a process called relay policy learning. This approach, introduced in [a paper prepublished on arXiv](#) and presented at the Conference on Robot Learning (CoRL) 2019 in Osaka, can be used to train artificial agents to tackle multi-stage and long-horizon tasks, such as object manipulation tasks that span over longer periods of time.

"Our research originated from many, mostly unsuccessful, experiments with very long tasks using [reinforcement learning](#) (RL)," Abhishek Gupta, one of the researchers who carried out the study, told TechXplore. "Today, RL in robotics is mostly applied in tasks that can be accomplished in a short span of time, such as grasping, pushing objects, walking forward, etc. While these applications have a lot value, our goal was to apply reinforcement learning to tasks that require multiple sub-objectives and operate on much longer timescales, such as setting a table or cleaning a kitchen."

Before they started developing their approach, Gupta and his colleagues reviewed previous literature to try and determine why longer tasks are particularly hard to tackle using current RL techniques. In their paper, they suggest that there are generally two main reasons for this.

First, it is hard for a robot to identify optimal solutions for solving long and complex tasks on its own. Second, it is difficult for the agent to successfully tackle a long [task](#) for which feedback is provided only at the end of a long sequence. Relay policy learning, the new approach to

learning that they presented, is designed to address both of these challenges head-on.



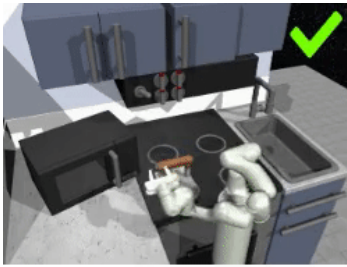
Credit: Gupta et al.

"To address the challenge of having robots solve long-horizon tasks on their own, we decided to simplify the problem and use human-provided demonstrations," Gupta said. "Solving long tasks is difficult because it's extremely hard to have a robot discover an interesting behavior on its own—human-provided demonstrations can be used as a guideline for interesting things to do in an environment."

The approach for [robot learning](#) proposed by Gupta and his colleagues has two distinct stages, one in which an agent learns by imitating humans and the other based on RL. In the imitation learning stage, a robot is fed human demonstrations of how to complete a task and produces goal-conditioned hierarchical policies.

In their study, the researchers used their approach to train an artificial agent called Franka on multi-stage and long-horizon manipulation tasks in a simulated kitchen environment, which was modeled using the physics simulator platform MuJoCo. This environment consisted of a kitchen with an openable microwave, four oven burners, an oven light

switch, a kettle, two hinged cabinets and a sliding cabinet door.



M, K, TB, L



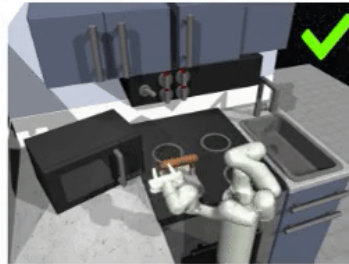
BB, TB, S, C



M, BB, S, C



M, K, S, C



M, BB, L, S



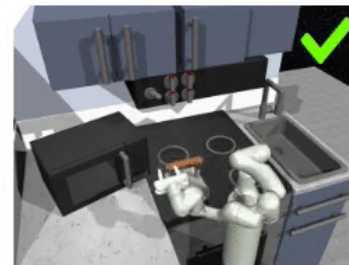
K, BB, TB, S



K, TB, L, S



M, L, S, C



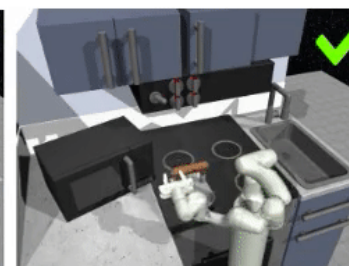
M, K, L, S



M, BB, TB, S



M, K, BB, C



M, BB, TB, C



M, K, BB, S

Key

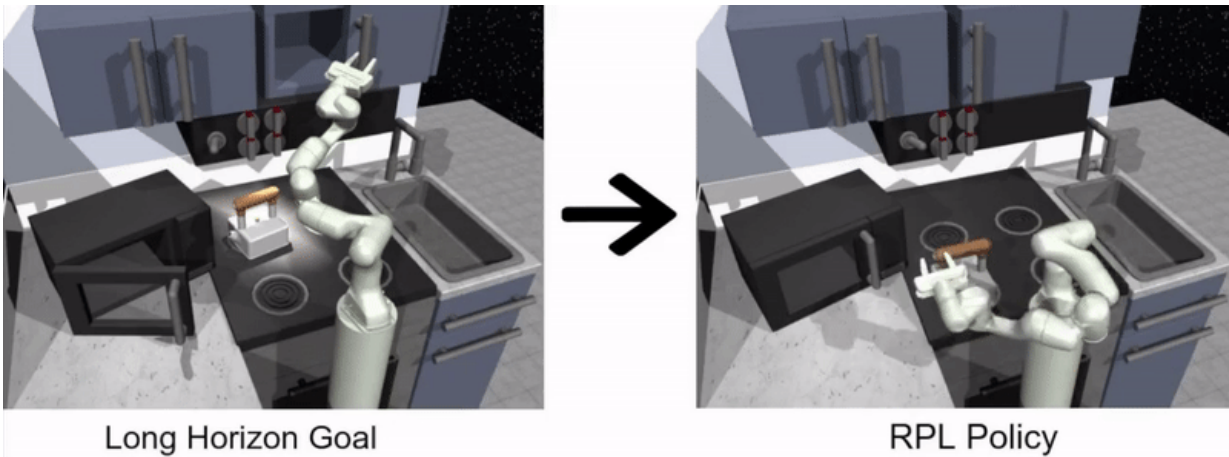
M = microwave
 K = Kettle
 BB = Bottom Burner
 TB = Top Burner
 L = Light
 S = Slider
 C = Hinge Cabinet

Credit: Gupta et al.

"Importantly, learning from demonstrations alone is not enough to solve the challenging tasks in our simulated kitchen environment," Karol Hausman, another researcher involved in the study, told TechXplore. "In order to improve upon this initial solution, we allow the robots to practice the tasks on their own to further refine their behaviors."

Essentially, using the relay policy learning method proposed by the researchers, an agent initially learns by processing human demonstrations of how to complete a given task and then continues learning on its own via RL. To make the process of learning long-horizon policies easier, the team used a new data-relabeling algorithm that allows an agent to learn goal-conditioned hierarchical policies.

"In order to tackle the challenge of sparse feedback, we use a [hierarchical structure](#) for our control policies: The high-level policy proposes goals that the low-level policy tries to accomplish—for example, close a cabinet, turn the burner off, etc.," Hausman explained. "This way, the task can be easily decomposed into smaller subproblems that can be solved with reinforcement learning bootstrapped from human-provided demonstrations."



Credit: Gupta et al.

Gupta, Hausman and their colleagues evaluated the effectiveness of relay policy learning for training robots in long-horizon tasks within the simulated kitchen environment they created, achieving very promising results. They found that with the right policy structure and demonstration data, their approach allowed robots to tackle much longer horizon tasks than they initially thought possible.

"We hope that our findings can open up new avenues of combining imitation and reinforcement learning research and gives us a potential direction that can allow robots to perform long, complex tasks," Hausman said.

In the future, the relay policy learning approach introduced by Gupta, Hausman and their colleagues could be used to train robots on a broader range of long-horizon tasks. The researchers have so far only tested their technique in a simulated environment; thus, it would be interesting to evaluate it in real-world settings and see whether it achieves equally promising results.

"As a next step, we would like to look into the problem of generalization beyond the demonstration data," Hausman said. "Eventually, we would also like to further improve the data-efficiency of our method, move to pixel observations and enable real-world learning on a physical [robot](#)."

More information: Relay policy learning: solving long-horizon tasks via imitation and reinforcement learning. arXiv:1910.11956 [cs.LG]. arxiv.org/abs/1910.11956

relay-policy-learning.github.io/

© 2019 Science X Network

Citation: Using imitation and reinforcement learning to tackle long-horizon robotic tasks (2019, November 13) retrieved 2 May 2024 from <https://techxplore.com/news/2019-11-imitation-tackle-long-horizon-robotic-tasks.html>

<p>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.</p>
--