https://ai.googleblog.com/2016/10/how-robots-can-acquire-new-skills-from.html

Google AI

Blog

The latest news from Google AI

How Robots Can Acquire New Skills from Their Shared Experience

Monday, October 3, 2016

Posted by Sergey Levine (Google Brain Team), Timothy Lillicrap (DeepMind), Mrinal Kalakrishnan (X)

The ability to learn from experience will likely be a key in enabling robots to help with complex real-world tasks, from assisting the elderly with chores and daily activities, to helping us in offices and hospitals, to performing jobs that are too dangerous or unpleasant for people. However, if each robot must learn its full repertoire of skills for these tasks only from its own experience, it could take far too long to acquire a rich enough range of behaviors to be useful. Could we bridge this gap by making it possible for robots to collectively learn from each other's experiences?

While machine learning algorithms have made great strides in natural language understanding and speech recognition, the kind of symbolic high-level reasoning that allows people to communicate complex concepts in words remains out of reach for machines. However, robots can instantaneously transmit their experience to other robots over the network - sometimes known as "cloud robotics" - and it is this ability that can let them learn from each other.

This is true even for seemingly simple low-level skills. Humans and animals excel at adaptive motor control that integrates their senses, reflexes, and muscles in a closely coordinated feedback loop. Robots still struggle with these basic skills in the real world, where the variability and complexity of the environment demands well-honed behaviors that are not easily fooled by distractors. If we enable robots to transmit their experiences to each other, could they learn to perform motion skills in close coordination with sensing in realistic environments?

We previously wrote about how multiple robots could pool their experiences to learn a grasping task. Here, we will discuss new experiments that we conducted to investigate three possible approaches for general-purpose skill learning across multiple robots: learning motion skills directly from experience, learning internal models of physics, and learning skills with human assistance. In all three cases, multiple robots shared their experiences to build a common model of the skill. The skills learned by the robots are still relatively simple -- pushing objects and opening doors -- but by learning such skills more quickly and efficiently through collective learning, robots might in the future acquire richer behavioral repertoires that could eventually make it possible for them to assist us in our daily lives.

Learning from raw experience with model-free reinforcement learning.

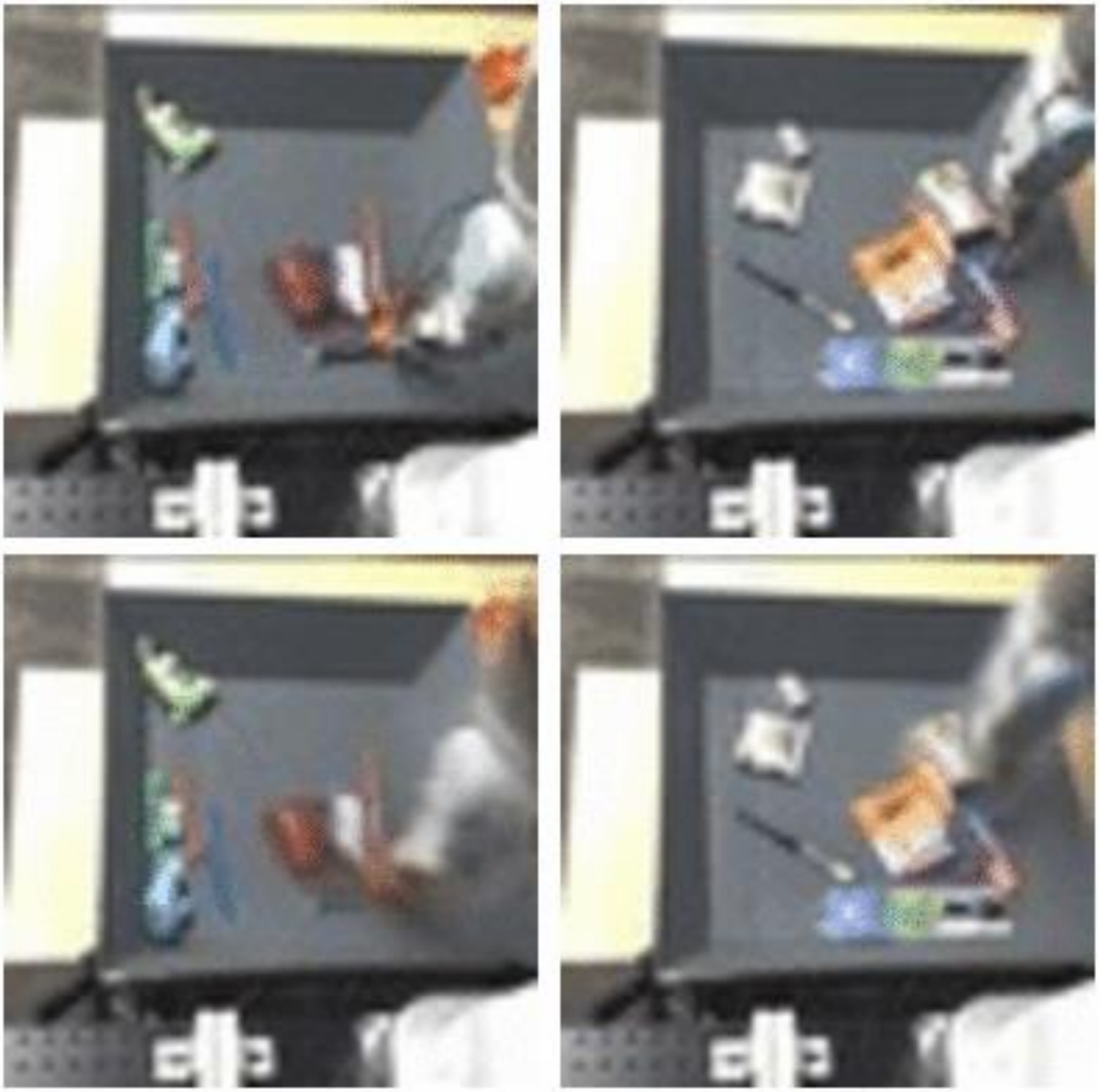
Perhaps one of the simplest ways for robots to teach each other is to pool information about their successes and failures in the world. Humans and animals acquire many skills by direct trial-and-error learning. During this kind of 'model-free' learning -- so called because there is no explicit model of the environment formed -- they explore variations on their existing behavior and then reinforce and exploit the variations that give bigger rewards. In combination with deep neural networks, model-free algorithms have recently proved to be surprisingly effective and have been key to successes with the Atari video game system and playing Go. Having multiple robots allows us to experiment with sharing experiences to speed up this kind of direct learning in the real world.

In these experiments we tasked robots with trying to move their arms to goal locations, or reaching to and opening a door. Each robot has a copy of a neural network that allows it to estimate the value of taking a given action in a given state. By querying this network, the robot can quickly decide what actions might be worth taking in the world. When a robot acts, we add noise to the actions it selects, so the resulting behavior is sometimes a bit better than previously observed, and sometimes a bit worse. This allows each robot to explore different ways of approaching a task. Records of the actions taken by the robots, their behaviors, and the final outcomes are sent back to a central server. The server collects the experiences from all of the robots and uses them to iteratively improve the neural network that estimates value for different states and actions. The model-free algorithms we employed look across both good and bad experiences and distill these into a new network that is better at understanding how action and success are related. Then, at regular intervals, each robot takes a copy of the updated network from the server and begins to act using the information in its new network. Given that this updated network is a bit better at estimating the true value of actions in the world, the robots will produce better behavior. This cycle can then be repeated to continue improving on the task. In the video below, a robot explores the door opening task.

With a few hours of practice, robots sharing their raw experience learn to make reaches to targets, and to open a door by making contact with the handle and pulling. In the case of door opening, the robots learn to deal with the complex physics of the contacts between the hook and the door handle without building an explicit model of the world, as can be seen in the example below:

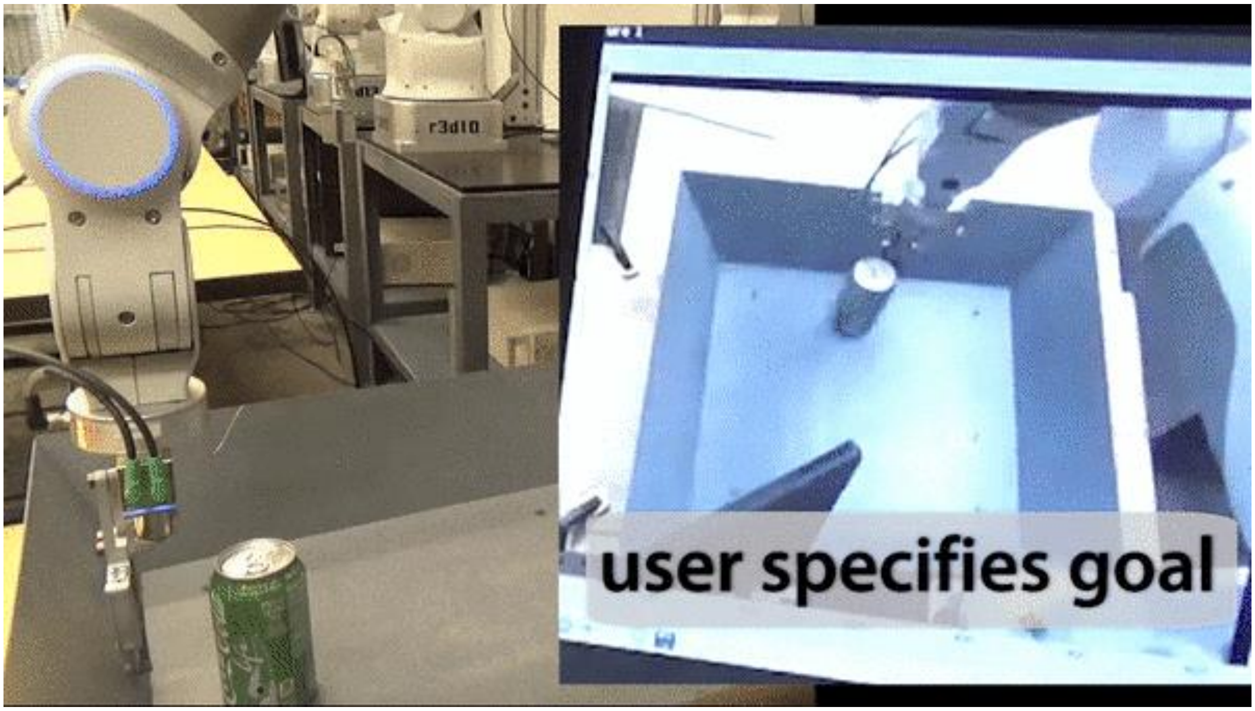
Learning how the world works by interacting with objects.
Direct trial-and-error reinforcement learning is a great way to learn individual skills. However, humans and animals don’t learn exclusively by trial and error. We also build mental models about our environment and imagine how the world might change in response to our actions.

We can start with the simplest of physical interactions, and have our robots learn the basics of cause and effect from reflecting on their own experiences. In this experiment, we had the robots play with a wide variety of common household objects by randomly prodding and pushing them inside a tabletop bin. The robots again shared their experiences with each other and together built a single predictive model that attempted to forecast what the world might look like in response to their actions. This predictive model can make simple, if slightly blurry, forecasts about future camera images when provided with the current image and a possible sequence of actions that the robot might execute:

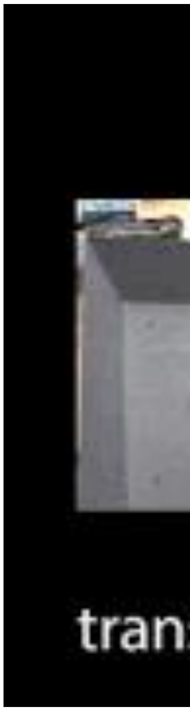


Top row: robotic arms interacting with common household items.
Bottom row: Predicted future camera images given an initial image and a sequence of actions.

Once this model is trained, the robots can use it to perform purposeful manipulations, for example based on user commands. In our prototype, a user can command the robot to move a particular object simply by clicking on that object, and then clicking on the point where the object should go:



The robots in this experiment were not told anything about objects or physics: they only see that the command requires a particular pixel to be moved to a particular place. However, because they have seen so many object interactions in their shared past experiences, they can forecast how particular actions will affect particular pixels. In order for such an implicit understanding of physics to emerge, the robots must be provided with a sufficient breadth of experience. This requires either a lot of time, or sharing the combined experiences of many robots. An extended video on this project may be found here.



Learning with the help of humans.

So far, we discussed how robots can learn entirely on their own. However, human guidance is important, not just for telling the robot what to do, but also for helping the robots along. We have a lot of intuition about how various manipulation skills can be performed, and it only seems natural that transferring this intuition to robots can help them learn these skills a lot faster. In the next experiment, we provided each robot with a different door, and guided each of them by hand to show how these doors can be opened. These demonstrations are encoded into a single combined strategy for all robots, called a policy. The policy is a deep neural network which converts camera images to robot actions, and is maintained on a central server. The following video shows the instructor demonstrating the door-opening skill to a robot:

Next, the robots collectively improve this policy through a trial-and-error learning process. Each robot attempts to open its own door using the latest available policy, with some added noise for exploration. These attempts allow each robot to plan a better strategy for opening the door the next time around, and improve the policy accordingly:

Not surprisingly, we find that robots learn more effectively if they are trained on a curriculum of tasks that are gradually increasing in difficulty. In our experiment, each robot starts off by practicing the door-opening skill on a specific position and orientation of the door that the instructor had previously shown it. As it gets better at performing the task, the instructor starts to alter the position and orientation of the door to be just a bit beyond the current capabilities of the policy, but not so difficult that it fails entirely. This allows the robots to gradually increase their skill level over time, and expands the range of situations they can handle. The combination of human-guidance with trial-and-error learning allowed the robots to collectively learn the skill of door-opening in just a couple of hours. Since the robots were trained on doors that look different from each other, the final policy succeeds on a door with a handle that none of the robots had seen before:

In all three of the experiments described above, the ability to communicate and exchange their experiences allows the robots to learn more quickly and effectively. This becomes particularly important when we combine robotic learning with deep learning, as is the case in all of the experiments discussed above. We’ve seen before that deep learning works best when provided with ample training data. For example, the popular ImageNet benchmark uses over 1.5 million labeled examples. While such a quantity of data is not impossible for a single robot to gather over a few years, it is much more efficient to gather the same volume of experience from multiple robots over the course of a few weeks. Besides faster learning times, this approach might benefit from the greater diversity of experience: a real-world deployment might involve multiple robots in different places and different settings, sharing heterogeneous, varied experiences to build a single highly generalizable representation.

Of course, the kinds of behaviors that robots today can learn are still quite limited. Even basic motion skills, such as picking up objects and opening doors, remain in the realm of cutting edge research. In all of these experiments, a human engineer is still needed to tell the robots what they should learn to do by specifying a detailed objective function. However, as algorithms improve and robots are deployed more widely, their ability to share and pool their experiences could be instrumental for enabling them to assist us in our daily lives.

The experiments on learning by trial-and-error were conducted by Shixiang (Shane) Gu and Ethan Holly from the Google Brain team, and Timothy Lillicrap from DeepMind. Work on learning predictive models was conducted by Chelsea Finn from the Google Brain team, and the research on learning from demonstration was conducted by Yevgen Chebotar, Ali Yahya, Adrian Li, and Mrinal Kalakrishnan from X. We would also like to acknowledge contributions by Peter Pastor, Gabriel Dulac-Arnold, and Jon Scholz. Articles about each of the experiments discussed in this blog post can be found below:

Deep Reinforcement Learning for Robotic Manipulation. *Shixiang Gu, Ethan Holly, Timothy Lillicrap, Sergey Levine.* [video]

Deep Visual Foresight for Planning Robot Motion. *Chelsea Finn, Sergey Levine.* [video]
[data]

Collective Robot Reinforcement Learning with Distributed Asynchronous Guided Policy Search.
Ali Yahya, Adrian Li, Mrinal Kalakrishnan, Yevgen Chebotar, Sergey Levine. [video]

Path Integral Guided Policy Search. *Yevgen Chebotar, Mrinal Kalakrishnan, Ali Yahya, Adrian Li, Stefan Schaal, Sergey Levine.* [video]