

Robots that understand contextual commands

August 31 2017, by Adam Conner-Simons



ComText allows robots to understand contextual commands such as, “Pick up the box I put down.” Credit: Tom Buehler/MIT CSAIL

Despite what you might see in movies, today's robots are still very limited in what they can do. They can be great for many repetitive tasks,

but their inability to understand the nuances of human language makes them mostly useless for more complicated requests.

For example, if you put a specific tool in a toolbox and ask a robot to "pick it up," it would be completely lost. Picking it up means being able to see and identify objects, understand commands, recognize that the "it" in question is the tool you put down, go back in time to remember the moment when you put down the tool, and distinguish the tool you put down from other ones of similar shapes and sizes.

Recently researchers from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) have gotten closer to making this type of request easier: In a new paper, they present an Alexa-like system that allows robots to understand a wide range of commands that require contextual knowledge about objects and their environments. They've dubbed the system "ComText," for "commands in context."

The toolbox situation above was among the types of tasks that ComText can handle. If you tell the system that "the tool I put down is my tool," it adds that fact to its knowledge base. You can then update the robot with more information about other objects and have it execute a range of tasks like picking up different sets of objects based on different commands.

"Where humans understand the world as a collection of objects and people and abstract concepts, machines view it as pixels, point-clouds, and 3-D maps generated from sensors," says CSAIL postdoc Rohan Paul, one of the lead authors of the paper. "This semantic gap means that, for robots to understand what we want them to do, they need a much richer representation of what we do and say."

The team tested ComText on Baxter, a two-armed humanoid robot developed for Rethink Robotics by former CSAIL director Rodney

Brooks.

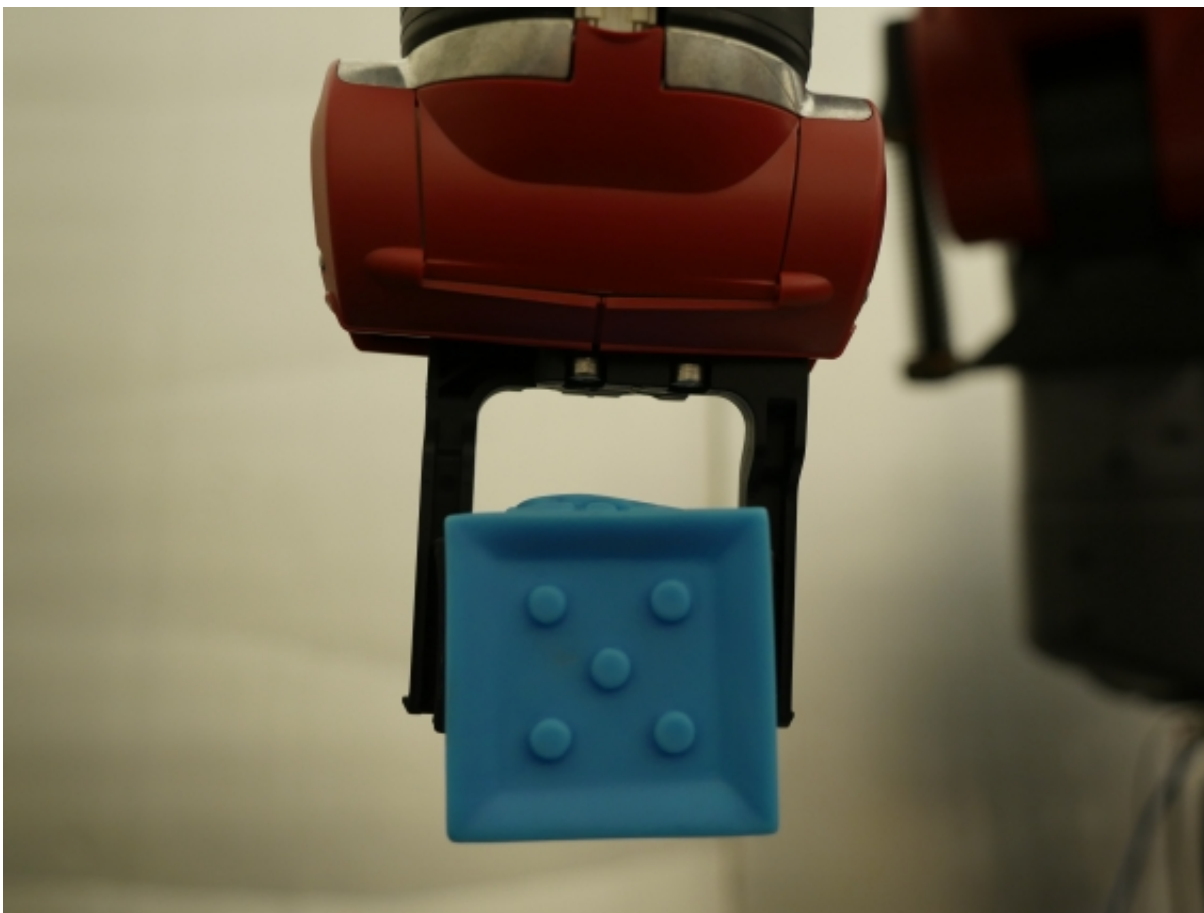
The project was co-led by research scientist Andrei Barbu, alongside research scientist Sue Felshin, senior research scientist Boris Katz, and Professor Nicholas Roy. They presented the paper at last week's International Joint Conference on Artificial Intelligence (IJCAI) in Australia.

How it works

Things like dates, birthdays, and facts are forms of "declarative [memory](#) ." There are two kinds of [declarative memory](#): semantic memory, which is based on general facts like the "sky is blue," and episodic memory, which is based on personal facts, like remembering what happened at a party.

Most approaches to robot learning have focused only on semantic memory, which obviously leaves a big knowledge gap about events or facts that may be relevant context for future actions. ComText, meanwhile, can observe a range of visuals and natural language to glean "episodic memory" about an object's size, shape, position, type and even if it belongs to somebody. From this [knowledge base](#), it can then reason, infer meaning and respond to commands.

"The main contribution is this idea that robots should have different kinds of memory, just like people," says Barbu. "We have the first mathematical formulation to address this issue, and we're exploring how these two types of memory play and work off of each other."



The Baxter robot picks up a block using the ComText system. Credit: Tom Buehler/MIT CSAIL

With ComText, Baxter was successful in executing the right command about 90 percent of the time. In the future, the team hopes to enable robots to understand more complicated information, such as multi-step commands, the intent of actions, and using properties about objects to interact with them more naturally.

For example, if you tell a robot that one box on a table has crackers, and one box has sugar, and then ask the robot to "pick up the snack," the hope is that the [robot](#) could deduce that sugar is a raw material and

therefore unlikely to be somebody's "snack."

By creating much less constrained interactions, this line of research could enable better communications for a range of robotic systems, from self-driving cars to household helpers.

"This work is a nice step towards building robots that can interact much more naturally with people," says Luke Zettlemoyer, an associate professor of computer science at the University of Washington who was not involved in the research. "In particular, it will help robots better understand the names that are used to identify objects in the world, and interpret instructions that use those names to better do what users ask."

More information: Rohan Paul et al. Temporal Grounding Graphs for Language Understanding with Accrued Visual-Linguistic Context, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (2017). [DOI: 10.24963/ijcai.2017/629](https://doi.org/10.24963/ijcai.2017/629)

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: Robots that understand contextual commands (2017, August 31) retrieved 23 April 2024 from <https://phys.org/news/2017-08-robots-contextual.html>

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.