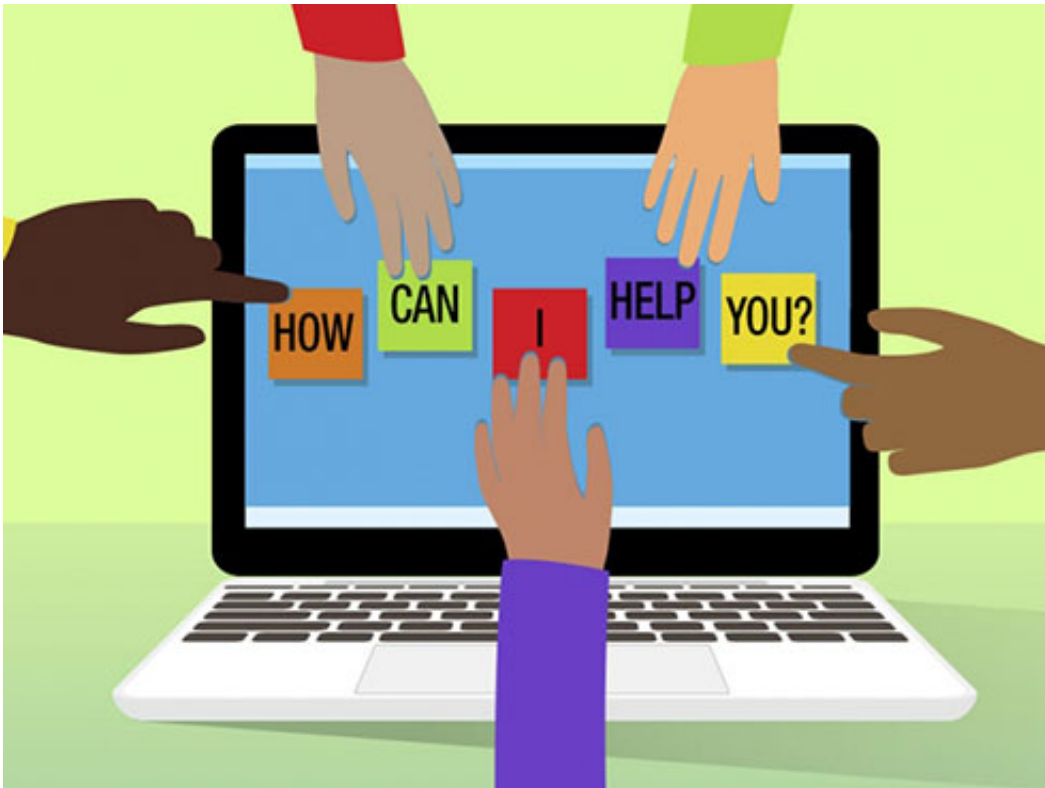


How Amazon's Mechanical Turk is becoming a research tool

September 14 2015, by Matt Windsor



This spring, Chris Callison-Burch, Ph.D., was in town to share an unusual approach to machine learning. This is one of the hottest topics in computer science: It is behind everything from Google's self-driving cars to Apple's Siri personal assistant.

Callison-Burch, an assistant professor at the University of Pennsylvania, is building a system that can automatically translate foreign languages into English—especially obscure dialects (from an American point of view) that can be of great interest to national security. He was in Birmingham at the invitation of Steven Bethard, Ph.D., a [machine learning](#) researcher and assistant professor in the UAB College of Arts and Sciences Department of Computer and Information Sciences.

In order to teach a computer to do something, Callison-Burch explained, you need to give it examples. Lots of examples. For a French-English translation, there are millions of sample texts available on the Internet. For Urdu, not so much.

Crowdsourced corpus

One way around this problem would be to pay professional translation services thousands of dollars to create the "corpus" of words you would need to train a computer to translate Urdu automatically. Callison-Burch has pioneered another approach: He paid some random folks on the Internet a few bucks at a time to do the work instead.

Callison-Burch is one of a growing number of researchers using Amazon Mechanical Turk, a service of the giant Internet company that bills itself as a "marketplace for work." Mechanical Turk, or MTurk, as it is known, "has almost become synonymous with crowdsourcing," Callison-Burch said. Anyone in need of help with a "human intelligence task" (Amazon's term) can post a job description, and the "reward" they are willing to pay. One recent afternoon, some of the 255,902 tasks available on MTurk included tagging photos on Instagram (4 cents per picture), typing out the text visible in distorted images (1 cent per image) and rating test questions for a biology exam for a researcher at Michigan State University (a penny per question—this is a popular price point).

Callison-Burch started out by giving Turkers and professional translators the same tasks. He encountered some trouble at first—respondents copying and pasting their assigned sentences into Google Translate, for example. "Quality control is a major challenge," Callison-Burch said. "It is important to design tasks to be simple and easy to understand."

So he tweaked his assignments to filter out people who weren't really native speakers, and added in some clever [quality control](#) mechanisms, such as getting additional Turkers to pick the best translations out of multiple versions of the same sentence. Callison-Burch was able to get remarkably close to the professional quality, for "approximately an order of magnitude cheaper than the cost of professional translation," he said.

Turk-powered translation could be particularly helpful in translating regional Arabic dialects, Callison-Burch noted. "Because standard machine translation systems are trained on written text, they don't handle spoken language well," he said. In a recent study, Callison-Burch and his collaborators found that "comments on Arabic newspaper websites were written in dialect forms about 50 percent of the time." A machine learning system trained in these dialects could offer vital clues about where a writer is from in the Middle East, for example, or about "his or her informal relationship with an interlocutor based on word choice."

Applications from obesity to philosophy

MTurk's brand of "artificial artificial intelligence" (Amazon's Turk tagline) could also be applied to other machine learning research at UAB, notes Steven Bethard. "Chris' work is fascinating," with applications from medicine to the social sciences, Bethard said.

UAB researchers are already putting MTurk to use. Andrew Brown, Ph.D., a research scientist in the Office of Energetics in the School of Public Health, has tested Turkers' ability to categorize biomedical

research studies. "We like to do some creative looks at what's been published and how," Brown said. For a recent paper, Brown and colleagues were interested in systematically evaluating nutrition-obesity studies. They wanted to find out whether studies with results that coincide with popular opinion are more likely to draw attention in the scientific community than studies that contradict the conventional wisdom. (They used citations as a proxy for the scientific community's opinion of a paper.)

The first step was to identify all the studies of interest. But "the problem is, there are 25 million papers in PubMed, and sometimes the keywords don't work very well," Brown said. "It helps to have a human set of eyes take a look at it." Instead of giving Ph.D.-level scientists the job, the researchers turned to MTurk. The Turkers successfully evaluated abstracts to identify appropriate studies and categorize the studied foods, then gathered citation counts for the studies in Google Scholar. (There was no significant link between public and scientific opinion when it came to the papers.)

"We found it to be useful," Brown said. "Expecting a perfect rating or an exhaustive rating from microworkers is probably a little premature, but on the other hand even trained scientists make mistakes." Brown plans to use crowdsourcing for future studies. "This is just one more tool to add to our research toolbox," he said.

Josh May, Ph.D., an assistant professor in the UAB College of Arts and Sciences Department of Philosophy, has been using MTurk for several years—asking Turkers to solve thorny moral dilemmas. "I present participants with hypothetical scenarios and ask them to provide their opinion about them—'Did the person act wrongly?'" May said. "Then I see whether responses change when the scenarios are slightly different, e.g., when a harm is brought about actively versus passively, or as a means to a goal versus a side effect. Statistical analysis can reveal

whether the differences are significant—providing evidence about whether the slight changes to the scenarios make a real difference in everyday moral reasoning."

Social justice and microwork

May, Brown and Callison-Burch share an interest in social justice for Turkers as well. "The main ethical issue with MTurk is exploitation," May said. "The going rate is often around a quarter for a few minutes of work, which typically adds up to less than the federal minimum wage, even when working quickly. This apparently isn't illegal given certain loopholes, but that doesn't make it moral. Just because someone will work for pennies doesn't mean we should withhold a living wage."

May's solution for his own research "is to estimate the time it will take most workers to complete the task and then pay them enough so that the rate would amount to at least minimum wage." Brown takes a similar approach—and when the Turkers work more slowly than expected, which drives down their overall wage, "there are bonus systems in place where you can give them something extra," he said.

Callison-Burch is using his programming skills to help Turkers earn fair wages. He has created a free browser extension (available at crowdworkers.com) that identifies high-paying jobs and makes it easier to identify job posters who have a large number of complaints.

Crowdsourcing operations such as MTurk represent an untapped resource for scientists of all stripes, Callison-Burch concluded. "Individual researchers now have access to their own data production companies," he said. "Now we can get the data we need to solve problems."

More information: "Using Crowdsourcing to Evaluate Published

Scientific Literature: Methods and Example." *PLoS ONE* 9(7): e100647.
[DOI: 10.1371/journal.pone.0100647](https://doi.org/10.1371/journal.pone.0100647)

Provided by University of Alabama at Birmingham

Citation: How Amazon's Mechanical Turk is becoming a research tool (2015, September 14)
retrieved 25 April 2024 from <https://phys.org/news/2015-09-amazon-mechanical-turk-tool.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.