DOI:10.1145/3068614

To view the accompanying paper, visit doi.acm.org/10.1145/3068663 rh

Technical Perspective
Humans and Computers Working Together on Hard Tasks

By Ed H. Chi

THE FIELD OF crowdsourcing and human computation has evolved considerably from its early days. At first, crowdsourcing was mainly conceived as a way to obtain ground truth labels for datasets, particularly image datasets, in the mid-2000s. Soon after, researchers began to utilize crowdsourcing for performing large-scale user studies of systems.[a,b] As our understanding of crowdsourcing continued to evolve, researchers realized the workers can be reserved ahead of time to perform real-time tasks.[c] Utilizing this idea, the system described in the following paper demonstrates how a crowd of workers can caption speech nearly as well as a professional captionist. Importantly, this paper was one of the first in a recent set of crowdsourcing papers that demonstrated how human workers can collaborate in concert with computing systems to accomplish a real-time task that is difficult for either one to do by itself. This is notable for many reasons, but let me first summarize the significance of this work.

First, the system demonstrated that significant innovation is needed to get human workers to productively perform the captioning task. For example, the Scribe system slows down the continuous speech for a brief period of time with the right volume changes to emphasize what passage to transcribe for the worker. The volume variations help with audio saliency. This technique is interesting to human-computer interaction (HCI) researchers, since it utilizes our intuition about how we can direct human attention, helping to transform individual untrained workers into better captionists.

Second, the system uses a MapReduce programming paradigm to divide and conquer the various pieces of the captioning tasks and coordinates the workers and their tasks through this organization paradigm. First introduced by Kittur et al.,[d] this is a clever application of the MapReduce paradigm, but instead of applying to computing tasks, the system applies the concept to organizing human tasks.

Third, impressively, to combine the partial contributions from individual workers, the system utilizes a sequence alignment algorithm to combine the streams of input from various workers. This is novel because most crowdsourcing systems use a simple majority voting approach to combine the worker inputs. The use of a sophisticated algorithm here is necessary to fit the captioning problem, and it points to the possibility of other combiner functions in other problems in future research. A natural extension of the alignment algorithm here would be to utilize a task-specific language model trained using deep learning.

From a historical perspective, augmenting humans has been at the very center of much personal computing and HCI research. There has been much talk about the degree in which machine learning (ML) will replace human labor (HL) in the future, but I think that is misguided. Instead, what we see in this research is a good example in which humans and machines work in concert on a very hard task that is currently still too difficult to do by either alone. Interestingly, this aligns well with a historical recounting of the code-breaking work by Turing and colleagues at Bletchley Park in a recent issue of *Communications*: "Another myth is that code-breaking machines eliminated human labor and code-breaking skill ... Technology transcended, rather than supplemented, human labor and bureaucracy."[e] The article points out the real challenge of the whole effort was a combination of the management of a (mostly female!) human operator force along with the Enigma machines. From my perspective, intelligent augmentation of our abilities is the real research frontier.

While we continue to explore the boundary of what is possible for machine intelligence, we should also be exploring the boundary of how humans will interact with machine intelligence. For example, how can we have an intelligent conversation with computing systems? Can I talk to a restaurant recommendation system while I drive home to get ready for a dinner date? How should my television respond if I say I wanted an exciting action film tonight that takes into account the tastes of other family members? If it doesn't have enough information on everyone in the room, will it (he/she?) ask intelligent questions while naturally conversing with my guests? Can I give feedback both via hand gestures as well as voice dialog?

Since an important application of machine intelligence is to augment humans in their desires, goals, and tasks, what we should do is to ask important research questions about human interactions with ML systems. In other words, we should have much better research of ML+HL, ML+HCI, and ML+Human Interaction, and this research is a shining example that points the way. C

a Kittur, A., Chi, E.H., Suh B.. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the ACM Conference on Human-Factors in Computing Systems*, ACM Press (Florence, Italy, 2008), 453–456.

b Egelman, S., Chi, E.H., Dow, S. Crowdsourcing in HCI research. *Ways of Knowing in HCI.* J.S. Olson and W.A. Kellogg, Eds. Springer, NY, 2014, 267–289.

c Bernstein, M., Brandt, J., Miller, R., and Karger, D. Crowds in two seconds: Enabling real-time crowd-powered interfaces. *UIST* 2011.

d Kittur. K, Smus. B., Khamkar. S., and Kraut. R.E. CrowdForge: Crowdsourcing complex work. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (2011), 43–52; http://dx.doi.org/10.1145/2047196.2047202

e Haigh, T. Colossal genius: Tutte, flowers, and a bad imitation of Turing. *Commun. ACM 60*, 1 (Jan. 2017), 29–35; https://doi.org/10.1145/3018994

Ed H. Chi is Research Lead Manager and Sr. Staff Research Scientist at Google Inc., Mountain View, CA.