Crowdsourcing and Crowd Work

Jeffrey P. Bigham
Carnegie Mellon University
jbigham@cs.cmu.edu

Walter S. Lasecki
University of Michigan
wslasecki@umich.edu

Chinmay Kulkarni
Carnegie Mellon University
chinmayk@cs.cmu.edu

Abstract

Crowdsourcing and human computation are useful in a number of real-world applications. Crowds generate large data sets useful for natural language processing and computer vision; they work together to formulate intelligent responses far beyond what we can automate; and they power intelligent interactive systems currently impossible with automated approaches alone. In this course, attendees will learn how to work with the crowd to enable research and practical applications. They will gain experience from the worker's perspective, receive an introduction to writing programs that work with existing sources of crowds, (*e.g.*, Amazon Mechanical Turk), apply usability principles for designing crowd tasks that elicit high-quality responses, use statistical methods to improve the quality of the work received, build systems that interface with crowd labor in real time, and conduct experiments to improve understanding of the differences between different sources of crowd work. The course will provide hands-on activities on each of these topics, and provide pointers to material that can provide more in depth information.

Detailed Description

Crowdsourcing underlies a surpisingly large portion of our interactions with intelligent systems [1]. Tools for search engines to content recommenation systems and digital personal assistants all commonly use human insight from large groups of workers. In this tutorial, we aim to make the use of crowdsourcing in a wide rage of appications more approachable to people with diverse specializations in both industry and academia. The hope is to broaden participation in crowdsourcing research, as well as get practitioners started off on the right foot in a domain that is often percieved as being a "dark art".

In this course, we will focus on foundational components of crowdsourcing and human computation, including:

- Application domains and impact

- History of human computation and crowdsourcing

- Platforms and considerations

- Workers and working conditions

- Tools for posting tasks

- Incentives and introductory game theory

- Workflows and human-in-the-loop algorithms

- Hybrid intelligence systems

- Applications

- Future of crowd work

Figure 1: The crowd person we developed for our crowdsourcing course at Carnegie Mellon University.

These topics will be organized into three units for the course: *(i)* being a crowd worker, *(ii)* crowdsourcing as a tool, and *(iii)* crowd-powered systems.

Being A Crowd Worker

The course will begin with time spent being a crowd worker on different platforms. Our experience teaching this material in a variety of other settings, namely a course at Carnegie Mellon University[1] and a course at the University of Michigan[2], has shown us that having students be crowd workers first both makes them more successful at crowdsourcing work later and instills empathy that may lead to more positive interactions with workers. One of the course instructors (Jeffrey Bigham)

---

[1]http://www.programthecrowd.com
[2]https://web.eecs.umich.edu/w̃lasecki/courses/hcs_fall2015/

wrote up his own experience working as a Mechanical Turk worker, which participants will be encouraged to read before participating in the course [3].

We will begin with a "getting to know each other" activity in which the class will be asked to collectively label a group photograph of themselves. This activity will ask them each to perform a microtask of labeling one person (not themselves). This has the benefit of getting people talking and allowing the instructors to get to know more of the participants. It also indirectly introduces many of the concepts that we will be covering in the course, such as worker coordination, quality control, etc.

We will then introduce Amazon Mechanical Turk (AMT) and ask workers to earn as much as they can in 20 minutes. To facilitate this process, we plan to email participants to ask them to sign up for AMT in advance of the course, given that it often takes days to be approved. The sign-up process itself is often illuminating for would-be workers because it is the first experience many will have with the very impersonal interactions that are the default with AMT. We will also have several existing AMT accounts available that participants can use in order to make progress as a worker.

Of course, AMT is not the only crowd platform, despite it being a popular one. We will also introduce participants to look at platforms that have other characteristics, such as UpWork[4], and consider how other labor pools could be usefully thought of as sources of crowd workers, *i.e.*, the crowd of participants at CHI and the crowd of students they have in their local or remote classes.

We will end this segment by introducing some of the tools and sites that workers use to earn more money on AMT and other crowdsourcing sites, and given students the opportunity to try again for 10 minutes to see if they were able to earn money more quickly. We will then ask the students to report back on their experiences.

Introduction to Crowdsourcing as a Tool

In this section, we will overview crowdsourcing as a tool for those who need to collect data and label data that they already have. This section will consist first of an overview presentation of the different issues that participants may want to consider: how to design high-quality tasks for workers to collect data (task design, gold standard question, input/output agreement), how to do quality control to ensure data integrity (statistcal techniques, decision-theoretic control [3]), and how to ethically treat worker participants in such tasks.

We will then walk paticipants through the process of designing a task, posting the task to AMT, collecting and analyzing data that comes back. Generally, marketplaces like AMT have fast enough response times that we have been able to get meaningful data back within only a short amount of time (within a single course session). As we wait for data to come back, we will talk about research that shows other ways of improving data quality, such as by having workers comment on each others' work, directly teach one another to do better work, and workflow design that helps ensure high-quality work.

---

[3]http://www.cs.cmu.edu/ jbigham/posts/2014/half-workday-as-turker.html

[4]http://www.upwork.com

Crowd-Powered Systems

The final section will consider how to use crowds as part of running interactive systems. This approach is incredibly useful for creating prototypes or even deployable systems for a variety of areas of HCI [7]. In this section, we will first introduce participants to work in this area, and then introduce the fundamental building blocks of real-time crowdsourcing, *e.g.*, real-time integration with crowdsourcing platforms [2], retainer models to gather groups quickly, and crowd agents as a pattern for balancing quality and quick responses [8]. Instructors will lead participants through an activity that will use Legion Tools [6, 4] to recruit a group quickly. They will also be encouraged to interact with the crowd via real-time chat using our deployed Chorus bot [9, 5] [5].

Summary and Outcomes

People will leave this workshop with a better understanding of crowdsourcing research in HCI and beyond, an understanding of the basic tools and techniques used in this space and the ability to use them, an intuition for the underpinning of designing effective crowd workflows, and knowledge of where to look for additional, topic-specific work in the area.

References

[1] Bigham, J. P., Bernstein, M., and Adar, E. Human-computer interaction and collective intelligence, 2015.

[2] Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., Miller, R., Tatarowicz, A., White, B., White, S., and Yeh, T. Vizwiz: Nearly real-time answers to visual questions. In *Proceedings of the 23rd User Interface Software and Technology*, UIST '10, ACM (New York, NY, USA, 2010), 333–342.

[3] Dai, P., Mausam, and Weld, D. S. Decision-theoretic control of crowd-sourced workflows. In *24th Association for the Advancement of Artificial Intelligence Conference* (2010).

[4] Gordon, M., Bigham, J. P., and Lasecki, W. S. Legiontools: a toolkit+ ui for recruiting and routing crowds to synchronous real-time tasks. In *Adjunct Proceedings of the 28th User Interface Software & Technology*, ACM (2015), 81–82.

[5] Huang, T.-H. K., Lasecki, W. S. L., Azaria, A. A., and Bigham, J. P. "is there anything else i can help you with?": Challenges in deploying an on-demand crowd-powered conversational agent. In *The Fourth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2016)* (2016).

[6] Lasecki, W. S., Gordon, M., Koutra, D., Jung, M. F., Dow, S. P., and Bigham, J. P. Glance: Rapidly coding behavioral video with the crowd. In *Proceedings of the 27th User Interface Software and Technology*, UIST '14, ACM (New York, NY, USA, 2014), 551–562.

[7] Lasecki, W. S., Homan, C., and Bigham, J. P. Architecting real-time crowd-powered systems. *Human Computation Journal* (September 2014).

[8] Lasecki, W. S., Murray, K. I., White, S., Miller, R. C., and Bigham, J. P. Real-time crowd control of existing interfaces. In *Proceedings of the 24th User Interface Software and Technology*, UIST '11, ACM (New York, NY, USA, 2011), 23–32.

[9] Lasecki, W. S., Wesley, R., Nichols, J., Kulkarni, A., Allen, J. F., and Bigham, J. P. Chorus: A crowd-powered conversational assistant. In *Proceedings of the 26th User Interface Software and Technology*, UIST '13, ACM (New York, NY, USA, 2013), 151–162.

---

[5]http://www.talkingtothecrowd.org