Tutorial: Making Better Use of the Crowd

Jennifer Wortman Vaughan
Microsoft Research
jenn@microsoft.com

1 Overview

Over the last decade, crowdsourcing has been used to harness the power of human computation to solve tasks that are notoriously difficult to solve with computers alone, such as determining whether or not an image contains a tree, rating the relevance of a website, or verifying the phone number of a business.

The natural language processing and machine learning communities were early to embrace crowdsourcing as a tool for quickly and inexpensively obtaining annotated data to train systems. Once this data is collected, it can be handed off to algorithms that learn to perform basic tasks such as translation or parsing.

Many times this handoff is where interaction with the crowd ends. The crowd provides the data, but the ultimate goal is to eventually take humans out of the loop. Are there better ways to make use of the crowd?

In this tutorial, I will begin with a showcase of innovative uses of crowdsourcing that go beyond data collection and annotation. I will discuss direct applications to natural language processing and machine learning, hybrid intelligence or "human in the loop" AI systems, and large scale studies of human behavior in online systems.

I will then dive into recent research aimed at understanding who crowdworkers are, how they behave, and what this can teach us about best practices for interacting with the crowd.

I will debunk the common myth that crowdsourcing platforms are riddled with bad actors out to scam requesters. In particular, I will describe the results of a research study that showed that crowdworkers on the whole are basically honest, and describe what requestors can do to encourage this honest behavior.

I will talk about experiments that have explored how to boost the quality and quantity of crowdwork by appealing to both well-designed monetary incentives (such as performance-based payments) and intrinsic sources of motivation (such as curiosity or a sense of doing meaningful work).

I will then discuss recent research—both qualitative and quantitative—that has opened up the black box of crowdsourcing to uncover that crowdworkers are not independent contractors, but rather a network of workers with a rich communication structure.

Taken as a whole, this research has a lot to teach us about how to most effectively interact with the crowd. Throughout the tutorial, I will discuss best practices for engaging with crowdworkers that are rarely mentioned in the literature but make a huge difference in whether or not your research studies succeed. (A few hints: Be respectful. Be responsive. Be clear.)

All material associated with the tutorial will be available at: http://www.jennwv.com/projects/crowdtutorial.html

2 Tutorial Outline

Part 1: The Potential of Crowdsourcing

- Direct applications to NLP and machine learning
- Hybrid intelligence systems
- Large-scale studies of human behavior online

Part 2: The Crowd is Made of People

- Crowdworker demographics
- Honesty of crowdworkers
- Monetary incentives
- Intrinsic motivation
- The network within the crowd
- Discussion and additional best practices

3   Instructor's Bio

Jenn Wortman Vaughan is a Senior Researcher at Microsoft Research, New York City, where she studies algorithmic economics, machine learning, and social computing, with a heavy focus on prediction markets and other forms of crowdsourcing. She is interested in developing general methods that allow us to reason formally about the performance of algorithms with human components in the same way that traditional computer science techniques allow us to formally reason about algorithms that run on machines alone. Jenn came to Microsoft in 2012 from UCLA, where she was an assistant professor in the computer science department. She completed her Ph.D. at the University of Pennsylvania in 2009, and subsequently spent a year as a Computing Innovation Fellow at Harvard. She is the recipient of Penn's 2009 Rubinoff dissertation award for innovative applications of computer technology, a National Science Foundation CAREER award, a Presidential Early Career Award for Scientists and Engineers (PECASE), and a handful of best paper or best student paper awards. In her "spare" time, Jenn is involved in a variety of efforts to provide support for women in computer science; most notably, she co-founded the Annual Workshop for Women in Machine Learning, which has been held each year since 2006.

References and Other Material

Please see the tutorial website: http://www.jennwv.com/projects/crowdtutorial.html