Human Computation and Crowdsourcing Systems

Walter S. Lasecki

EECS 598, Fall 2015
Who am I?

- http://wslasecki.com

- New to UMich!
  - Prof in CSE, SI
  - BS, Virginia Tech, CS/Math
  - PhD, University of Rochester, CS… half while at CMU

- Taught an undergrad variant of this course at CMU

- Research: HCI / AI
  - Human computation for intelligent systems
  - Real-time and continuous crowdsourcing
  - Underlying models and limitations
  - Access technology for people with disabilities
  - Interested in research on these topics? Let me know!
My Research
Who am I?

- http://wslasecki.com
- New to UMich!
  - Prof in CSE, SI
  - BS, Virginia Tech, CS/Math
  - PhD, University of Rochester, CS… half while at CMU
- Taught an undergrad variant of this course at CMU
- Research: HCI / AI
  - Human computation for intelligent systems
  - Real-time and continuous crowdsourcing
  - Underlying models and limitations
  - Access technology for people with disabilities
  - Interested in research on these topics? Let me know!
Definitions

(Warning: terms not always well-defined)
What is “Human Computation”? 

- Integrating people into computational process

- People fill a well-defined functional role
  - In contrast to most human endeavors which is not well defined
  - Allows automated systems to work with human input

- Often structured output, but not always
  - Labels, confidence values, image boundaries, ordered text
  - Design feedback,

- “Human Computation”
  - Luis von Ahn, PhD thesis title, CMU
Why Human Computation?

- Artificial Intelligence (AI) cannot currently solve everything
- Even problems that are automatable are not always solved
- Example: ESP Game
  - von Ahn et al., CHI 2004
  - Label images with pairs of people
Historical Examples of Human Computation

- Human computation is not “new”
  - “When Computers Were Human” — David Alan Grier
  - Works Progress Administration
  - Needed to give people jobs, and find ways to make it useful
  - Used non-expert to compute canonical mathematical tables
What is “Crowdsourcing”? 

- An open call to a group of people

- “Crowdsourcing”
  - “Crowdsourcing is the act of taking a job traditionally performed by a designated agent ... and outsourcing it to...a large group of people in the form of an open call.”
  - [ Jeff Howe, Wired ]

- Books
  - Jeff Howe: Crowdsourcing
  - James Surowiecki: The Wisdom of Crowds
Why Crowdsourcing?

- No one worker will always be available

- Open call allows for more available human intelligence
  - Allow for the creation of on-demand systems
  - Even real-time becomes possible — 1s responses or less with multiplexing

- Any individual has a chance of error
  - With groups of workers, we might be able to reduce this error rate
  - Especially for ephemeral workers

- Collectively, we can get pieces that work together in parallel
Types of Crowds (At a Glance)

- **User crowds**
  - User/community generated content, interaction traces
  - Social media, online forums
  - Game players (e.g., GWAP)

- **Paid crowds**
  - Amazon Mechanical Turk
  - Focus groups
  - Contracting platforms (expert)

- **Volunteer crowds**
  - Community-sourcing
  - Activism / demonstrations

- … (continued next class)
Historical Examples of Crowdsourcing

- **Crowdsourcing is not “new”**
  - Generalization: collective intelligence
  - Not restricted to people: emergent behaviors

- **Guessing the weight of an Ox [James Surowiecki]**
  - Average of a group was within 1% of the correct answer
  - Group performs better than any expert
Wait, what’s the difference?

[Quinn & Bederson, CHI 2011]
Wait, what’s the difference?

“Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call.” [24]
Wait, what’s the difference?

“...a technique that makes use of human abilities for computation to solve problems.” [8,74]

“A computational process that involves humans in certain steps...” [73]

[Quinn & Bederson, CHI 2011]
Wait, what’s the difference?

“... applications and services that facilitate collective action and social interaction online with rich exchange of multimedia information and evolution of aggregate knowledge...” [48]

“... the interplay between persons' social behaviors and their interactions with computing technologies” [15]
Wait, what’s the difference?

Groups that collectively act with intelligence (including phenomena like emergent behavior)

[Quinn & Bederson, CHI 2011]
Wait, what’s the difference?

Find insight into large sets of data, such as datasets generated by collective systems

[Quinn & Bederson, CHI 2011]
Connections: Computer Science

- Human-Computer Interaction (HCI)
- Artificial Intelligence (AI) / Machine Learning (ML)
- Multi-Agent Systems
  - Economics / [Algorithmic] Game Theory / Incentive Mechanism Design
- Parallel Computing
- [[discuss]]
Connections: Models of Work

- Firms
  - Management science
  - “Value of a firm” \(\Rightarrow\) efficiency of organization

- Work processes
  - Adam Smith: division of labor
  - Assembly lines, unit productions
  - Taylorism: ‘scientific management’ of workers
  - Modern evidence-based process management

- Group dynamics
  - Team structure / cools
  - Organizational behavior/psychology
Connections: Consensus-Finding

- Voting theory / election systems
  - Find collective answers
  - Avoid / leverage manipulation

- Collaboration in teams
  - Structures
  - Workflows / organization

- Collective intelligence
  - Emergent behavior
Connection: Social Sciences

- Social networks / communities

- Cognitive science
  - Individual and group cognition

- Psychology
  - Interpersonal behaviors
  - Biases
  - Incentives (esp. intrinsic)

- Political science

- ... [[discuss]]
What is possible?

- **Previous:**
  - Processes for creating knowledgebases / producing answers

- **Now:**
  - Systems to label images/audio/etc. — often, output can train AI
  - Real-time / on-demand systems (few-second response latency, no down time)
  - Groups of non-experts can outperform experts
  - Less-restricted, more creative tasks

- **Future:**
  - Complex open-ended tasks that result in computationally usable answers
  - Millisecond-level latency even with human assistance
  - Online training of AI/ML systems
  - … You tell me.
Dangers of these models?

“Magnus, Robot Fighter #1” (Feb. 1, 1963)
[ Course Info ]

web.eecs.umich.edu/~wlasecki/courses/hcs_fall2015/
Course Objectives

- Introduce you to the field of crowdsourcing
  - And to some of the prior work on crowd-powered systems

- Show you [some of] the open problems in crowdsourcing research
  - And give you a sense of why they matter

- Gain experience working with crowdsourcing tools and platforms

- Contribute some novel work in this space
Course Focus

- **System building**
  - That does not mean you need to be an expert system builder! (but it doesn’t hurt)
  - Theory, qualitative studies, data analysis through the lens of practical systems

- **Literature**
  - Readings will cover classic work in crowdsourcing systems
  - Additional reading material may be provided periodically as a supplement

- **Projects**
  - Start with small/medium sized to get used to the tools and platforms
  - Single large ‘final’ project — ideally something novel enough to consider publishing
  - Proposed deal: the more novel the project, the more support available (to make it fair to all)
Your challenge
(should you choose to accept it)
Do crowdsourcing research ... as a crowd.
[ Survey ]

shoutkey.com/real
[ Logistics ]
Schedule

- **T/Th: 12PM to 1:30PM**
  - Feel free to bring [non-disruptive] lunch
  - Short lecture, in-class collaboration + discussion session
  - Teams present papers; meet about the coming week’s papers
  - Presenter teams and discussant teams

- **Discussion section**
  - *No discussion section this week*
  - Usage? ...Let’s crowdsource a decision!
    - [Vote]: presentations or separate presentation review?
Present in class
- Spend 30min of class presenting and discussing
- Teams prepare outside of class, meet with me

Present separately
- More in-class collaboration
- Present in discussion sec.
- Spend some time at the end of class meeting about slides
Other Course Information

- In addition to the research readings, I’ll try to post occasional re-cap writings

- Office hours: *via email, or by appointment*

- Course website: Coming soon.

- Canvas? ...TBD

- Crowdsourcing design of this course — your feedback helps!
  - Consider this an open call =)
Next Class

- Lecture:
  - Crowds, types, and platforms

- Readings:
  - “Human Computation: A Survey and Taxonomy of a Growing Field”
  - “Demographics of Mechanical Turk”
  - Discuss as a group in class
  - Group assignments for future papers

- **Project 1: “Be a crowd worker”**
  - Details / assignment