From needs to ideas

MICHAEL BERNSTEIN
Today

- Synthesizing your needfinding results
- From needs to ideas
- Milestone 3
- Overview of crowdsourcing research
Synthesizing your needfinding results

:thumbs_up:
Workers: summary

• In order to make money, workers need to find tasks that result in payment and maintain their reputation.

• They want to find reputable requesters that communicate with workers, create clear HITs, and justify rejected work.
Find reliable requesters

- Feedback: Workers need to know what they’re doing wrong so that they don’t make the same mistakes again. They also need to make sure that they weren’t rejected unfairly. ~Team SneakyLittleHobbitses  Team Dubs
- Communication: need to understand their work quality quickly. ~@vaastav
- Trust: need to trust/compare/rate requesters (especially new ones).  ~@shreygupta98
Money & Reputation

- Prioritizing HITs: Workers need to decide if the HITs are worth their time.
  ~Team SneakyLittleHobbitses @rajashri92, @natashahervatta

- Getting paid for time: “Workers deserve to be paid for the work regardless of the fact if it is up to the Requester's standards or not and they deserve better value for their work.” ~@vaastav

- Find work that can make you $: Workers need to be able to search for tasks that match their interests and
Money & Reputation

- Find work that can make you $: Workers need to be able to search for tasks that match their interests and abilities. ~@shreygupta98
- e.g., Some workers aren’t good/fast with survey tasks while some are.
- To maintain qualifications (reputation) to complete HITs and make money ~@vaastav
  - Laura, only works tasks as long as she is sure that any new task wouldn't get her approval rating below 99%. Rochelle is hesitant working with new requesters.
Money & Reputation

- Do good work: Workers need to do a task well. ~@shreygupta98
Requesters: summary

- Ultimately, requesters want to easily create HITs, to have their HITs be accepted by workers, and to receive good quality HIT results.

- …all as seamlessly and quickly as possible.
  - “Magic moment” of near-instant results (@spamgirl)
Authoring

- Need to make something that gets the right results, the first time.
- Peter doesn't use in-built templates as he feels they are not good enough for his HITs.
Quality

• Ensure output is good, original and not copied ~@yoni.dayan

• Chris (from the panel) talks about how writing clear instructions are important for the requesters to get the desired output.
Quality

• Need to post clear, unambiguous requirements  ~@alipta
• It's quite difficult for requesters to know whether…
  • 1) a worker tried hard but the question was unclear or very difficult or an edge case, or
  • 2) a worker wasn't really putting in a best effort.  
    ~@freddiev
Quality

- “Requesters want better [faster] ways to ensure the validity of the work.” ~@vaastav

- Design tasks to avoid turkers gaming the system
  ~@vaastav
Reputation & communication

• The Requester’s reputation (on online forums) depends on the number of unfair rejections they make, so they try not to make harsh decisions.
  ~Team SneakyLittleHobbiteses

• [Need to] … improve their relationship with workers.
  ~Team duka

• It's very difficult to scale engagement with workers if they are more and more of them, and it's impossible to answer to all the emails.
  ~@yoni.dayan
Michael’s synthesis

• There are two main factors at play: trust and power.
  • How do I trust who you say you are? How do I trust that the results I get are results that will be good? How do I trust that you’ll respect me as a worker, and pay me accordingly?
  • Who has the power to post work? To edit other peoples’ posted work? To return results to the requester? Can I, as a worker, send it back myself, or does someone else need to vet it?
Brainstorming

Because it seems like a black art.
flare  focus  flare  focus
“How might we...?” questions

- Turn large needs into actionable charges
  - e.g., “How might we make crowdrater feedback feel more like trusted, safe spaces?”
- A useful way to ground a brainstorm
The Goldilocks of How Might We

- A good “How Might We…” question is:
  - Not so broad that it is inapproachable
    *How might we help people organize all their digital media?*
  - Not so narrow that it suggests a solution
    *How might we help people retrieve their favorite digital media with just a click?*
  - In a happy middle ground:
    *How might we help weekend extreme sports enthusiasts organize their digital media?*
Dark horse ideas

• Include one idea and prototype that is intentionally far out or nearly impossible. Sometimes, it wins.
• Even when it doesn’t win, it helps prevent design fixation.
Crowdsourcing research

To prompt your ideas…
Can the whole be greater than the sum of the parts?

- Can technology guide large groups of people to tackle bigger, harder problems than they could in isolation?
- Help large groups come together to act…
  - At an expert level,
  - On complex tasks,
  - At a high level of quality.
Early crowdsourcing research

[Little et al., HCOMP 2009]

Two distributed workers work independently, and a third verifier adjudicates their responses

You (misspelled) (several) (words). Please spellcheck your work next time. I also notice a few grammatical mistakes. Overall your writing style is a bit too phoney. You do make some good (points), but they got lost amidst the (writing). (signature)
Early crowdsourcing research

[Grier 2007]

Two distributed workers work independently, and a third verifier adjudicates their responses

1760

British Nautical Almanac
Neil Maskelyne
In answer to your letter of the 23. instant
I compute diurnal motion of the sun as follows.
Assume the radius of the earth's orbit to be 1.
Assume the distance of the sun from the earth to be 1.
The sun's apparent motion is then the difference between the two.

\[ \text{Apparent motion} = \sqrt{r^2 - 1} \]

where \( r \) is the radius of the earth's orbit.

The letter formula must be used when \( r \) is large or near 100, but may be used safely in all cases.

Example:

\[ \text{Apparent motion} = \sqrt{100^2 - 1} \]

The result is approximately 99.99.

If the result is greater than 100
\[ \text{Apparent motion} = 100 \]

This is the case for the planets, excepting Venus.

The planets are only set down to minutes, but a

\[ \text{Apparent motion} = 100 \]

is the closest we can go.

The planets are

- Mercury
- Venus
- Earth
- Mars
- Jupiter
- Saturn
- Uranus
- Neptune
- Pluto

The positions of the planets are

\[ \text{Position} = \text{Apparent motion} \]
Two people doing the same task in the same way will make the same errors.
Mathematical Tables Project

- WPA project, begun 1938
- Calculated tables of mathematical functions
- Employed 450 human computers
- The origin of the term computer
Etymology

• Crowdsourcing term coined by Jeff Howe, 2006 in Wired

• “Taking [...] a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call.”
Success: games with a purpose

Label every image on the internet using a game

[von Ahn and Dabbish, CHI ’06]
Success: scientific collaboration

- FoldIt: protein-folding game
- Amateur scientists have found protein configurations that eluded scientists for years
More successes

Largest encyclopedia in history

Disaster reporting

Kasparov vs. the world

Collaborative math proofs

NASA Clickworkers

DARPA Red Balloon Challenge
Paid Crowdsourcing

- Pay small amounts of money for short tasks
- Amazon Mechanical Turk: Roughly five million tasks completed per year at 1-5¢ each [Ipeirotis 2010]
- Population: 40% U.S., 40% India, 20% elsewhere
- Gender, education and income are close mirrors of overall population distributions [Ross 2010]

<table>
<thead>
<tr>
<th>Task</th>
<th>Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label an image</td>
<td>$0.02</td>
</tr>
<tr>
<td>Transcribe audio clip</td>
<td>$0.05</td>
</tr>
</tbody>
</table>
Major topics of research

Crowd algorithms
[Little et al., HCOMP 2009]

Incentives and Quality
[Mason and Watts, HCOMP 2009]
[Dow et al., CSCW 2012]

Crowd-powered systems
[Bernstein et al., UIST 2010]
[Bigham et al., UIST 2010]

AI for HCOMP
[Dai, Mausam & Weld, AAAI 2010]

Complex Work
[Kittur et al., UIST 2011]
Crowdsourcing algorithms
Goal: guide crowds as they work

- Designing crowdsourcing algorithms is often like designing a user interface that will keep a user “in bounds” on your application
- Challenges
  - Taking unexpected action
  - Trying too hard
  - Trying not hard enough
Crowdsourcing algorithm

- A generalized version of a workflow

- Iterative algorithms [Little et al. 2009]
  - Hand off from one worker to the next

- Most crowdsourcing processes are more parallel, but less interesting algorithmically
Crowdsourcing algorithms

- Open-ended editing: Find-Fix-Verify [Bernstein et al., UIST ’10]
- Graph search [Parameswaran et al., VLDB ’11]
- Clustering [Chilton et al., CHI ’13]
- and many more...

- When write an algorithm?
  If you tried this in a straightforward way, would crowds fail? Why?
Incentives and quality
Incentives

• Does paying more produce better work?
  • More work, but not higher-quality work
    [Mason and Watts, HCOMP ’09]

• Does feedback produce better work?
  • Self-assessment and expert assessment both improve the quality of work
    [Dow, Kulkarni, Klemmer and Hartmann, CSCW ’11]
Incentives
[Shaw, Horton and Chen, CSCW ’11]

- Which of these approaches improve quality?
  - Comparison to other workers
  - Normative claims: “it’s important that you try hard”
  - Solidarity: your team gets a bonus if you are right
  - Humanization: “thanks for working; I’m Aaron.”
  - Reward or punish accuracy with money
  - Reward or punish agreement with money
  - Bayesian truth serum: predict others’ responses
  - Bet payment on the accuracy of your responses
Incentives
[Shaw, Horton and Chen, CSCW ’11]

- Which of these approaches improve quality?
  - Comparison to other workers
  - Normative claims: “it’s important that you try hard”
  - Solidarity: your team gets a bonus if you are right
  - Humanization: “thanks for working; I’m Aaron.”
  - Reward or punish accuracy with money
  - Reward or punish agreement with money
  - Bayesian truth serum: predict others’ responses
  - Bet payment on the accuracy of your responses
Motivations
[Antin and Shaw, CHI ’12]

- Ask workers: “I am motivated to do HITs on Mechanical Turk...”
  - To kill time
  - To make extra money
  - For fun
  - Because it gives me a sense of purpose

- List experiment: vary which reasons appear in the list, and ask how many reasons the participant agrees with
  - This technique counters social desirability bias
Motivations

[Antin and Shaw, CHI ’12]

- US workers
  - 40% overreporting of money as a reason to work
- India-based workers
  - 142% underreporting of killing time and 60% underreporting fun as reasons
  - Money was not over- or under-reported
Communitysourcing

Engaging Local Crowds to Perform Expert Work Via Physical Kiosks

Kurtis Heimerl, Brian Gawalt, Kuang Chen
Tapan Parikh, Björn Hartmann
University of California, Berkeley

Hacking motivation

CHI 2012
Judging quality explicitly

- **Gold standard judgments** [Le et al., SIGIR CSE ’10]
  - Include questions with known answers
  - Performance on these “gold standard” questions is used to filter work
- **Get Another Label** [Sheng, Provost, Ipeirotis, KDD ’08]
  - Estimate the correct answer and worker quality jointly
Judging quality implicitly

[Rzeszotarski and Kittur, UIST ’12]

- Observe low-level behaviors
  - Clicks
  - Backspaces
  - Scrolling
  - Timing delays
- SVMs on these behaviors predict work quality
- Limitation: models must be built for each task
Crowd-powered systems
Why do it?

- Embed crowd intelligence inside of user interfaces and applications we use today
Automatic clustering generally helps separate different kinds of records that need to be edited differently, but it isn’t perfect. Sometimes it creates more clusters than needed, because the differences in structure aren’t important to the user’s particular editing task. For example, if the user only needs to edit near the end of each line, then differences at the start of the line are largely irrelevant, and it isn’t necessary to split based on those differences. Conversely, sometimes the clustering isn’t fine enough, leaving heterogeneous clusters that must be edited one line at a time. One solution to this problem would be to let the user rearrange the clusters manually, perhaps using drag-and-drop to merge and split clusters. Clustering and selection generalization would also be improved by recognizing common text structure like URLs, filenames, email addresses, dates, times, etc.
VizWiz
[Bigham et al., UIST ’10]

• Visual question answering for the blind

• 1 to 2 minute responses by keeping workers on fake tasks until needed
Crowd-powered databases

- Database with open-world assumptions:
  SELECT * FROM ice_cream_flavors
- Several university flavors
  - Berkeley: CrowdDB [Franklin et al., SIGMOD ’11]
  - MIT: Qurk [Marcus et al., CIDR ’11]
  - Stanford: Deco [Parameswaran et al. ’11]
- Tackling many important optimization questions: e.g., joins, ranking, sorting
Realtime crowdsourcing

[Bernstein et al., UIST '11]
Realtime crowdsourcing

- Realtime captioning using shotgun gene sequencing techniques

![System Overview Diagram]

- Scribe
  - System Overview
  - Speech Source
    - we have a crystal that has a two-fold axis...
  - Flash Media Server
  - Caption Stream
    - we have a crystal that has a two-fold axis...
  - Crowd Corrections
  - Output
    - we have a crystal that has a two-fold axis...
  - Merging Server
    - has a two-fold axis
    - we have a crystal
    - have a crystal that has
Artificial intelligence for crowds
TurKontrol: AIs guiding crowds
[Dai, Mausam and Weld, AAAI ’10]

• Workflow planning as a decision-theoretic optimization problem
• Trade off quality vs. number of workers required
  • POMDP to decide: do we need a vote? do we need more voters? do we need more improvement?
Complex work
CrowdForge

[Kittur et al., UIST ’11]

- Crowdsourcing as a map-reduce process
- To write a wikipedia page, partition on topics, map to find facts and then reduce into a paragraph
Turkomatic
[Kulkarni, Can, and Hartmann, CSCW ’12]

- Let the workers decide on task design
- Is a task too complicated for $D$? If so, ask for sub-tasks and recurse. If not, do it yourself.

- Creating a blog with content:
Careers in crowd work

[Kittur et al., 2013]

- More and more people are engaging in online paid work: programmers, singers, designers, artists, …
- Would you feel comfortable with your best friend, or your own child, becoming a full-time crowd worker?
- How could we get to that point? What would it take?
  - Education
  - Career advancement
  - Reputation