From needs to ideas

MICHAEL BERNSTEIN
Today

- Synthesizing your needfinding results
- From needs to ideas
- Milestone 3
- Overview of crowdsourcing research
Prof. Andrew Ng
Hangout

Stanford artificial intelligence and deep learning pioneer
Chat with him about research more generally at 9am PDT on 3/26
The role of our peer feedback

It's easy to tear things down, but that's not the point here. The goal of feedback should be to identify the amazing ideas to share.
Synthesizing your needfinding results

Again, amazing work!
Worker needs

- Workers need to be able to quickly find tasks they'd want to work on
  - Monday evening panel workers cite the challenge in identifying or distinguishing tasks because of poor tagging
  - Reddit discussion: exorbitant amount of time spent trying to find tasks and do the mental calculations to find the opportunities that match them best (e.g. time to complete the task on average, average $ per minute on task, requirements to complete the task)
  - Interpretation: Finding a task that matches well with the worker's skillset and pays well takes a significant amount of (unpaid) time.
Worker needs

- Workers need to feel they are being fairly compensated for their work.
  - Reddit discussion cites that the payment system for HITs is not adaptive and does not take into account changing marketplace conditions (supply/demand) and pricing of tasks based on those changes.
  - Interpretation: Monetary compensation is the primary motivator for many crowd workers.
Worker needs

- Workers need to feel like they are treated fairly and respectfully, and have a voice in the platform.
- Comment on Turkopticon: "Got a mass rejection from some hits I did for them! Talked to other turkers that I know in real life and the same thing happened to them. Their rejection comments are also really demeaning. Definitely avoid!"
- Interpretation: Unreasonable rejections and low payments lead workers to feel disrespected. The implicit assumption on MTurk is that workers are unskilled and replaceable. They can do little if their work is rejected.
Worker needs

- Workers need to be able to expose their skills so they can get work they are qualified for and advance their skills.
  - Monday evening panel workers from oDesk cite that most employers will not work with them until they have enough feedback or past work on the platform
  - Interpretation: If users cannot get new work without feedback, this makes it difficult for new users to establish their reputation and get jobs that will help develop their skillsets.
Requester needs

- Requesters need to get their HITs completed (quickly / correctly).
  - Requester asking on forum why nobody is doing his HITs (7-minute, 25-cent surveys - a very low wage)
  - Interpretation: Requesters want their HITs done, and when nobody's doing them, they do not know the reason why (e.g. it is because he is underpaying workers).
Requester needs

- Requesters need to be able to trust the results they get.
  - Requesters will often rely on previous workers whose results they can trust, and add mechanisms to detect spammers, or manually verify some results.
  - Interpretation: If spammers are not caught, this brings the correctness of results into question. If requesters are not sure the results are correct, they may need to discard the data.
Requester needs

- Requesters need to have workers who have the appropriate skills and demographics do their tasks.
  - Requesters worry that they are not able to verify self-reported demographics for surveys.
  - Interpretation: Workers' self-reported skills and demographics are often not viewed as trustworthy. This is a problem for surveys, which need to have correct demographic data to be useful.
Requester needs

• Requesters need to be able to easily generate good tasks.
  • Companies hire full-time developers to deal with the complexities of posting microtasks on MTurk. Requesters often develop their own tools and workflow systems on top of Amazon's.
  • Interpretation: The process of authoring HITs is currently difficult and makes crowd-work inaccessible to potential requesters.
Requester needs

- Requesters need to be able to price their tasks appropriately.
  - Requesters asking on forums about the appropriate amount they should pay for their HITs.
  - Interpretation: Requesters often don't have a good intuition of what the appropriate wage for their task would be in terms of price per HIT.
Requester needs

- Requesters need workers to trust them.
  - Requesters say they are reluctant to reject work, because they fear they might get bad reviews.
  - Interpretation: Workers are more likely to do HITs if the requester seems trustworthy. Requesters do not want bad reviews, because they may result in workers ignoring the requester's HITs.
Requester needs

• Requesters need workers to trust them.
  • Requesters say they are reluctant to reject work, because they fear they might get bad reviews.
  • Interpretation: Workers are more likely to do HITs if the requester seems trustworthy. Requesters do not want bad reviews, because they may result in workers ignoring the requester's HITs.
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.
“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?
Cardinal rules of brainstorming

- Go for quantity
- Defer judgment
- Enforce a time limit
- Riff and build on others’ ideas
Roles in brainstorming

- Facilitator
  - Not the manager or leader!
  - Keeps the conversation flowing
- Recorder
  - Gets everything up on the whiteboard (or sketchbook)
- Participant
How might we...?
One tactic: Powers of Ten

- When you’re considering a question, scale it up and scale it down
- Ice skating as a deliberate controlled vulnerability…
  - Could 1000 people at once engage in a controlled vulnerability activity? Could someone do it alone?
  - Could we do this in extreme heat environments?
Design space axes

- Pull out high-level dimensions along which your observations or ideas vary, then look for holes.

Diagram:

- **Invulnerable**
  - Laser tag
  - Paintball
  - Watching an action movie
- **Vulnerable**
  - Ice skating
  - Trust fall
  - ?
- **Solo**
- **Group**
Warm-up time

- Your goal: loosen people up so that they aren’t self-conscious about sharing ideas with the group
- Reinforce practices of volume, riffing, and fluidity
- Improv games work quite well
Dark horse prototyping

- 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.
Milestone 3

Time to launch into brainstorming!
Basic outline

• Work with your team to brainstorm as many ideas as you can under two headings: trust and power
  • “How might we” enable workers to trust the requester’s intention to pay?”
  • “How might we” enable requesters to trust the results they get back?
• Ten ideas per heading — they can be just sketches at this point. The result of your brainstorming.
Step two: elaboration

- Pick two ideas per heading and elaborate them so that your peer researchers could easily understand them.
  - Sketches? Longer descriptions? Storyboards?
- And finally, add one more dark horse idea for a total of three
- These three ideas are the ones you most want your peers to hear about
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

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

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

1760

British Nautical Almanac
Neil Maskelyne
Work distributed via mail
Two people doing the same task in the same way will make the same errors.

Charles Babbage
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]

<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>

- Population: 40% U.S., 40% India, 20% elsewhere
- Gender, education and income are close mirrors of overall population distributions [Ross 2010]
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 clustering 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
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
Conflict and coordination

- What happens to collaboration costs as Wikipedia grows? [Kittur, Suh, Pendleton, and Chi, CHI '07]

Amount of direct work on articles goes down, and activity on coordination pages goes up
Conflict and coordination

- As more editors join, which kinds of coordination techniques succeed? [Kittur and Kraut, CSCW ’08]
  - Explicit: participation in talk pages
  - Implicit: set direction by making edits

More editors only improves article quality only with implicit coordination — a few take on a disproportionate amount of 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