Ideas

MICHAEL BERNSTEIN
Today

- Themes from the brainstorming exercise
- Milestone 4
- Overview of crowdsourcing research
Reminder: Andrew Ng

Stanford artificial intelligence and deep learning pioneer
Chat with him about research more generally at 9am PDT on 3/26
How’s the new feedback site?

Do you like it better? Is the feedback more constructive in nature?
Insights in your ideas

Once again, blown away by the variety and depth.
Task clarity

• How might workers+requesters work together help produce higher-quality task descriptions?

• Ideas:
  • New tasks go to workers who improve the task before it goes live
  • Voting on hit design
  • New tasks go into a holding pattern where they need to be "voted in" by workers
  • Task templates (can we do better than AMT’s templates?)
  • Stronger categorization of work
  • Artificial turker that has to understand tasks before they go live
Data and Results

- How might workers+requesters work together to produce higher-quality results?

- Ideas:
  - Workers review other workers’ work (like on MobileWorks)
  - Could we create a crowd contractor who is in charge of each submission getting good results?
Disputes

• How might workers+requesters work together to produce higher-quality results?
• Ideas:
  • Moderators can review if rejected work should be disputed (similar idea here)
  • Dispute resolution process
  • Bad actors get hellbanned to a separate circle for a time being, or their tasks done more slowly
  • Force requesters to mediate previous problems before they can post new tasks moving forward
Empathy

- How might we build more empathy between workers and requesters?

- Ideas:
  - Meetings outside of the system to build trust
  - Make workers act as requesters, and visa versa
  - Requesters can send gifts to workers
  - A certain amount of requester costs are reserved to pay bonuses to workers
  - Humanizing worker profiles
Transparency

• How might we make payment clear and transparent?

• Ideas:
  • Standardize task pricing
  • Required minimum wage? if you stay above the 15th percentile of “good work throughput”
  • Checkpoints where you get reviewed and paid after every N tasks
  • Offer increased compensation to the first few people who take the task
Reputation

• How might we make payment clear and transparent?
• Ideas:
  • Leveling up as a worker and requester: gives you better wages (Michael's thought there), first availability of tasks for workers, etc. Also worker levels
  • Ranking top workers and making them most available
  • Workers rate requesters as they do tasks, used to learn matching for other workers.
  • Tasks are first available to workers who match according to skill and performance
Milestone 4

Flare and focus!
Pick three themes

- As a team, pick three of the six themes that you’re going to focus on with your brainstorming this week
- Read the linked ideas on the wiki for each theme
  - …and let us know if we missed good ones
- Ask yourselves:
  - What’s similar about the ideas in this theme?
  - What’s different about the ideas in this theme?
For each of the three themes...

- Use the others’ ideas to launch a brainstorm on the theme
- Once you have lots of ideas, pick one to explore in more detail
  - Sketches, descriptions, storyboards
Next…

• We’ll start collapsing these ideas into concrete directions to prototype
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.
Work distributed via mail
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
- Kasparov vs. the world
- NASA Clickworkers
- Disaster reporting
- Collaborative math proofs
- 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
- Population is broad and replicates many well-known studies

**Label an image**
Reward: $0.02

**Transcribe audio clip**
Reward: $0.05
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]
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
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:
  \[\text{SELECT} \ast \text{FROM} \text{ice\_cream\_flavors}\]
- 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

**Scribe System Overview**

- Speech Source: we have a crystal that has a two-fold axis...
- Flash Media Server
- Speech
- Caption Stream
- Crowd Corrections
- Output: we have a crystal that has a two-fold axis
- Merging Server

Output: 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
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