Crowdsourcing

MICHAEL BERNSTEIN CS 376



sign up for progress meetings

Can the whole be greater than the sum of the parts?

- 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.

Can technology guide large groups of people to tackle bigger,



Early crowdsourcing research [Little et al., HCOMP 2009]

adjudicates their responses

You would dell sourced pool Real grade and not inter a de alter with a for gamentich mother Queral your writing Agels is a die to pool give to the to

Two distributed workers work independently, and a third verifier



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 adjudicates their responses



1760 British Nautical Almanac Neil Maskelyne

Two distributed workers work independently, and a third verifier



in answer to your totter of the 23. instant bompite D' dist. from a star by loganthmes, thus I the D, Sthe star, I Da for fundrular great circle. Let fall from I to PL, perfondiular to it in T. There suppose to be to be than go. By Log. c, \$++, \$\$=+, \$2 & AL-OD= LD c, df = -c, dD + c, df + c, df~= c, + s, PS - s, PD+ c, 2D The latter formula much be used when PD is large, or near go; but may be use da fely inall cased. loample P. 20-30 - C - - 9. 9715876 PS 89-58-30 t-13.3857585 1 25.25.30 -- C -- 9.9557589 --19 ---- s. 9.9999879 \$\$ 89.29.36-. + 12.0534829---- 8. co-ar. ---- 0.0000170 ---- c - - - 9.9544269 Jal 85.0.0 Lob -- H. 29.36 Ja 25- A7.30 If P should be greater than go then Alt AD = Dasin the sockere As the ilanets place's, excepting the Work distributed via mail

is donger. Is danter. grand. stanolles] at Noon. al Noon ? of the 2 of the Jay S. D. M.S. D. M.S. 1 2 20 0 51 2 58 291 2 3 2 45 83 48 32 3 3 15 46 23 4 28 29 3 4 3 29 3 54 4 54 28 4 Whica m. 5 4 12 35 47 5 4 34 5 M 6 4 26 19 33 4 57 26 6 02 7 5 10 1231 4 32 51 7 5 24 12 11 3 51 51 8 antarca 6 8 16 36 2 56 35 9. 1 Regular 10 6 22 24 16 1 50 19 10 11 11.7 6 34 4 0 37 6NG11 & aquila O Ja Vinca mp 127 20 44 43 0 38 25 112 Vd. 13 8 4 54 27 1 51 21 13 Formalh 148 19 0 452 57 1 14 Ja 159 3 0 17 3 51 19 15 Ja 14 Val antazer 16 9 16 49 18 4 31 12 16 24 17:10 0 24 84 54 54 54 17 a arecti 13 10 13 41 50 5 1 59 18 M 19 10 26 40 41 4 53 8 19 14 2011 9 20 28 4 29 54 20 Vd 21 11 21 42 30 3 54 21 21 220 3 49 35 3 8 48 22 aldebad 23 0 15 45 40 2 15 32 23 netio





Charles Babbage

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 leff 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."





CROWD SOURCINC

OWER OF THE FUTURE OF BUS











Success: games with a purpose Label every image on the internet using a game

[von Ahn and Dabbish, CHI '06]



Success: scientific collaboration

- Foldlt: proteinfolding game
- Amateur scientists have found protein configurations that eluded scientists for years



| | Rank: 17 48: Pro Peptide Group Competition | Score: 9 |
|--|---|----------|
| | # Group Name 1 The Lone Folder 2 Street Smarts 3 Illinois 4 Berkelev | |
| | Player Competition 16 psen 17 kathleen 18 versat82 19 darktorres 20 ccarrico 21 mbjorkegren 22 sslickerson Chat | 9092 |



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
- at I-5¢ each [lpeirotis 2010]

Label an image

Reward: \$0.02

Transcribe audio clip

Reward: \$0.05

- Population: 40% U.S., 40% India, 20% elsewhere
- distributions [Ross 2010]

Amazon Mechanical Turk: Roughly five million tasks completed per year

Gender, education and income are close mirrors of overall population

13

Major topics of research

I also which a for garnetist mother Dural is a bit ton plang. In to which god pill,



Automatic clustering generally helps separate different kinds of ecords that need to be edited differently, but it isn't perfect. mes it creates more clusters than needed, because the rences in structure aren't important to the user's particular editing sk. For example, if the user only needs to edit near the end of each then differences at the start of the line are largely irrelevant, and sn't necessary to split based on those differences. Conversely ometimes the clustering isn't fine enough, leaving heterogeneous isters that must be edited one line at a time. One solution to this blem would be to let the user rearrange the clustering manually haps 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

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 relevant to a specific task. Conversely, sometimes the clustering isn't fine enough, leaving heterogeneous clusters that must be edited one line at a time. solution to this problem would be to let the user rearrange the clustering manually using drag-and-drop edits. Clustering and selection generalization would also be improved by recognizing common text structure like URLs, filenames, email addresses, dates times, etc.



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]

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

Complex Work [Kittur et al., UIST 2011]

14

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



16

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?

18

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?
 - work
 - [Dow, Kulkarni, Klemmer and Hartmann, CSCW '11]

Self-assessment and expert assessment both improve the quality of



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]

- - 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

Ask workers: "I am motivated to do HITs on Mechanical Turk..."



Motivations [Antin and Shaw, CHI '12]

• US workers

- 40% overreporting of money as a reason to work
- India-based workers
 - reasons
 - Money was not over- or under-reported

142% underreporting of killing time and 60% underreporting fun as



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

25

Judging quality explicitly

- Gold standard judgments [Le et al., SIGIR CSE '10]
 - Include questions with known answers
- Get Another Label [Sheng, Provost, Ipeirotis, KDD '08]
 - Estimate the correct answer and worker quality jointly
 - Try it! https://github.com/ipeirotis/Get-Another-Label

• Performance on these "gold standard" questions is used to filter work



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



Interfa@vizard of Turk Wizard of Oz



Soylent

Page: 1 of 1 Words: 138 🕉

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.

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 structural differences aren't important to the user's editing task. 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 using drag-and-drop edits. Clustering and selection generalization would also be improved by recognizing common text structure like URLs, filenames, email addresses, dates, times, etc.

JIII

Shortn

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, as structure differences aren't important to the editing task. Converselye, if the 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 using drag-and-drop edits. Clustering and selection generalization would also be improved by recognizing common text structure like URLs, filenames, email addresses, dates, times, etc.

🗐 🛱 🗟 🗿 🗮 170% 😑 ──



VizWiz [Bigham et al., UIST '10]

Visual question answering for the blind

What color is this pillow?

What denomination is this bill?

Do you see picnic tables What temperature is my across the parking lot?



(89s) l can't tell. (105s) multiple shades of soft green, blue and gold



(24s) 20 (29s) 20



(13s) no (46s) no

I to 2 minute responses by keeping workers on fake tasks until needed

oven set to?

(69s) it looks like 425 degrees but the image is difficult to see. (84s) 400 (122s) 450

Can you please tell me

what this can is?

(183s) chickpeas. (514s) beans (552s) Goya Beans

What kind of drink does this can hold?



(91s) Energy (99s) no can in the picture (247s) energy drink



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 'II]
 - Stanford: Deco [Parameswaran et al. '11]
- Tackling many important optimization questions: e.g., joins, ranking, sorting





Realtime crowdsourcing



Realtime captioning using shotgun gene sequencing techniques

intelligence for CROWGS

TurKontrol: Als 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
 - POMDP to decide: do we need we need more improvement?





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



Ask most people who plan to travel to New York City what they want to see while they are there and invariably you will hear about the top tourist attractions: the Empire State Building, the Statue of Liberty, and the Grand Central Terminal, with the Empire State Building probably coming in as number one on the list of "must see" for visitors to the city. No wonder: the Empire State Building has a long history, having celebrated its seventy-fifth anniversary on May 1, 2006. Yet the Statue of Liberty is also a popular tourist destination.



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 subtasks 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

