Crowdsourcing: Wisdom of Crowds

Fall 2013

Dongwon Lee, Ph.D.
“Collective intelligence can be brought to bear on a wide variety of problems, and complexity is no bar... conditions that are necessary for the crowd to be wise: diversity, independence, and ... decentralization”
"Crowdsourcing represents the act of a company or institution 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. … The crucial prerequisite is the use of the open call format and the large network of potential laborers..."
“Crowdsourcing as an online, distributed problem-solving and production model that leverages the collective intelligence of online communities to serve specific organizational goals”
“Human computation is simply computation that is carried out by humans… Crowdsourcing can be considered a method or a tool that human computation systems can use…”

By Edith Law & Luis von Ahn
Eg, Francis Galton, 1906

Weight-judging competition:
1,197 (mean of 787 crowds) vs. 1,198 pounds (actual measurement)
Eg, StolenSidekick, 2006

- A woman lost a cellphone in a taxi
- A 16-year-old girl ended up having the phone
  - Refused to return the phone
- Evan Guttman, the woman’s friend, sets up a blog site about the incident
  - http://stolensidekick.blogspot.com/
  - http://www.evanwashere.com/StolenSidekick/
  - Attracted a growing amount of attention → the story appeared in Digg main page → NY Times and CNN coverage → Crowds pressure on police …
- NYPD arrested the girl and re-possessed the phone

http://www.nytimes.com/2006/06/21/nyregion/21sidekick.html?_r=0
Eg, “Who Wants to be a Millionaire”

Asking the audience usually works ➔ Audience members have diverse knowledge that can be coordinated to provide a correct answer in sum.
Eg, DARPA Network Challenge, 2009

- To locate 10 red balloons in arbitrary locations of US
- Winner gets $40K
- MIT team won the rate with the strategy:
  - 2K per balloon to the first person, A, to send the correct coordinates
  - 1K to the person, B, who invited A
  - 0.5K to the person, C, who invited B, …
Eg, Threadless.com

- Sells t-shirts, designed/voted by crowds
- Artists whose designs are chosen get paid
Eg, reCAPTCHA

As of 2012

Captcha: 200M every day

ReCaptcha: 750M to date
Crowdfunding, started in 2009

- Project creators choose a deadline and a minimum funding goal
  - Creators only from US, UK, and Canada
- Donors pledge money to support projects, in exchange of non-monetary values
  - Eg, t-shirt, thank-u-note, dinner with creators
  - Donors can be from anywhere
- Eg, Pebble, smartwatch
  - 68K people pledged 10M
Crowdsourcing Landscape Beta v2

Marketplaces
- Elance
- Guru
- Upwork
- Freelancer
- oDesk
- CrowdSpring
- TopCoder
- Kaggle
- Zooppa
- IdeaScale
- Qmarkets
- IdeaManagement
- IdeaExchange
- IdeaScale
- IdeaSpring
- IdeaPool

Platforms
- Data.com
- Aardvark
- Mahalo
- ChaCha
- Google
- Quora
- 40Billion
- Sockgame
- SEER
- ODesk
- CrowdSonic
- CrowdSpring
- IdeaSpring
- IdeaPool
- IdeaScale
- Qmarkets
- IdeaManagement
- IdeaExchange
- IdeaPool
- IdeaScale
- Qmarkets

Crowd Services
- Crowdflower
- CrowdProcess
- CrowdDesign
- CrowdVentures
- CrowdPlatforms
- CrowdPlatforms
- CrowdPlatforms

Content and Product Markets
- Data
- Content
- Product
- SaaS
- Cloud
- IoT
- Mobile
- Blockchain
- Virtual
- Reality

Non-profit
- Causes
- Citizen Engagement
- Contribution
- Non-profit
- Crowdfund
- Crowdfund
- Crowdfund
- Crowdfund
- Crowdfund

Crowd Crowdfunding
- 40Billion
- Sockgame
- SEER
- ODesk
- CrowdSonic
- CrowdSpring
- IdeaSpring
- IdeaPool
- IdeaScale
- Qmarkets
- IdeaManagement
- IdeaExchange
- IdeaPool
- IdeaScale
- Qmarkets

Excerpted from
Getting Results From Crowds by Ross Dawson and Steve Byghall
For definitions, analysis, free book chapters, and other crowdsourcing resources go to:
www.resultsfromcrowds.com

Note: examples only; see website for full list of crowdsourcing services

http://www.resultsfromcrowds.com/features/crowdsourcing-landscape/
What is Crowdsourcing?

- Many definitions
- A few characteristics
  - Online and distributed
  - Open call & right incentive
  - Diversity and independence
  - Top-down & bottom-up

Q: What are the **computational** issues in crowdsourcing?
- **Micro-tasks** for large crowds
What is Computational Crowdsourcing?

- Focus on computational aspect of crowdsourcing
- Mainly use micro-tasks
- Algorithmic aspect
- Optimization problem with three parameters

When to use Computational Crowdsourcing?
- Machine can’t do the task well
- Large crowds can do it well
- Task can be split to many micro-tasks
Human Computation

- Human Computation
  - Design a solution using both automated computers and human computers
  - Maybe a closed set of workers
  - Crowdsourcing greatly facilitates human computation
- Social computing
  - Social behavior
- Collective intelligence
  - May not human

By G. Li

Alexander J. Quinn, Benjamin B. Bederson: Human computation: a survey and taxonomy of a growing field. CHI 2011:1403-1412
## Tasks for Crowdsourcing

<table>
<thead>
<tr>
<th>Task Granularity</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex Tasks</td>
<td>• Build a website</td>
</tr>
<tr>
<td></td>
<td>• Develop a software system</td>
</tr>
<tr>
<td></td>
<td>• Overthrow a government?</td>
</tr>
<tr>
<td>Simple Projects</td>
<td>• Design a logo and visual identity</td>
</tr>
<tr>
<td></td>
<td>• Write a term paper</td>
</tr>
<tr>
<td>Macro Tasks</td>
<td>• Write a restaurant review</td>
</tr>
<tr>
<td></td>
<td>• Test a new website feature</td>
</tr>
<tr>
<td></td>
<td>• Identify a galaxy</td>
</tr>
<tr>
<td>Micro Tasks</td>
<td>• Label an image</td>
</tr>
<tr>
<td></td>
<td>• Verify an address</td>
</tr>
<tr>
<td></td>
<td>• Simple entity resolution</td>
</tr>
</tbody>
</table>

By G. Li
Computational Crowdsourcing

- **Requesters**
  - People submit some tasks
  - Pay **rewards** to workers

- **Marketplaces**
  - Provide crowds with tasks

- **Crowds**
  - Workers perform tasks

---

**Find an outlier among three images**

<table>
<thead>
<tr>
<th>Submit tasks</th>
<th>Collect answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon MT</td>
<td>Crowdflower</td>
</tr>
<tr>
<td>CloudCrowd</td>
<td></td>
</tr>
</tbody>
</table>

**Find tasks**

**Return answers**

1. Submit tasks
2. Collect answers
3. Find tasks
4. Return answers
Crowdsourcing Marketplaces

• Platforms for posting/performing (often micro) tasks

• Those who want to have tasks done via crowdsourcing ➔ Requesters
  • Eg, companies, researchers

• Those who want to perform tasks for monetary profits ➔ Workers
  • Eg, individuals for extra income
Crowdsourcing Platforms

- Notables ones:
  - Mechanical Turk (AMT)
  - CrowdFlower
  - CloudCrowd
  - Clickworker
  - SamaSource
AMT: mturk.com
AMT (cont)

- **Workers**
  - Register w. credit account (only US workers can register as of 2013)
  - Bid to do tasks for earning money
- **Requesters**
  - First deposit money to account
  - Post tasks
    - Task can specify a qualification for workers
  - Gather results
  - Pay to workers if results are satisfactory
AMT (cont)

- Tasks
  - Called **HIT** (Human Intelligence Task)
  - Micro-task

- Eg
  - Data cleaning
  - Tagging / labeling
  - Sentiment analysis
  - Categorization
  - Surveying
  - Photo moderation
  - Transcription

Translation task
### AMT: HIT List

- **Workers qualification**
  - Location is not VN
  - Location is not TR
  - Location is not RO
  - Location is not PK
  - Location is not PH
  - Location is not IN
  - Location is not ID
  - Location is not HK
  - HIT approval rate (%) is greater than 96
Can You Find the Provided Phone Number or Street Address on this Website?

Overview
In this task, you'll be provided a web page for a business, including its name, address, and phone number. Your goal is to answer a few questions about the business on the web page.

IMPORTANT: Sometimes the business will have multiple locations, and you will have to search the website for the specific business that we provide in order to verify the website.

Step by step instructions:
- Click the link to go to the provided site.
- First, please tell us whether or not the name of the business on the provided website is a close or identical match to the name of the business shown at the top of the page.
- Next, please tell us whether the provided business has the correct address.
- Please be sure to click the appropriate option if the site information is the same.

Wrinkles Day Spa
Phone: +61893455333
Street: Shop 5a Stirling Central Shopping Centre, 478 Wanneroo Rd
City: Westminster
State: WA
Postalcode (Zip): 6061
Country Code: AU

Click here to visit the website.

Is the name of the business on the web page similar or Identical to 'Wrinkles Day Spa'?
- Yes: the name of the business is similar to Wrinkles Day Spa
- Yes: the name of the business is nearly identical to Wrinkles Day Spa
- No: the name is very different from Wrinkles Day Spa

For the first option, the street number does not need to match, just the street, Shop 5a Stirling Central Shopping Centre, 478 Wanneroo Rd.
Factors for Crowd Algorithms

- Latency (or execution time)
  - Worker pool size
  - Job attractiveness

- Monetary cost
  - Number of questions
  - Cost per question

- Quality of answers
  - Worker maliciousness
  - Worker skills
  - Task difficulty

How much $$ does we spend?
How long do we wait for?
How much is the quality of answers satisfied?
Eg, Text Transcription

- **Problem**: one person can’t do a good transcription
- **Key idea**: iterative improvement by many workers

*Greg Little et al.* “Exploring iterative and parallel human computation processes.” HCOMP 2010
Eg, Text Transcription (cont)

- Please improve the transcription of this handwriting.
- People will vote whether to approve your changes.

You (?) (?) (?) (work). (?) (?) (?) work (not) (time). I (?) (?) a few grammatical mistakes. Overall your writing style is a bit too (phony). You do (?) have good (points), but they got lost amidst the (writing). (signature)

improvement $0.05

By R. Miller
Eg, Text Transcription (cont)

Please select the better transcription for this handwriting. Differences are highlighted in yellow.

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

You (?) (?) (?) work. (?) (?) (?) work (not) (time). I (?) (?) a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (?) have good (points), but they got lost amidst the (writing). (signature)

By R. Miller
“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)"

According to our ground truth, the highlighted words should be “flowery”, “get”, “verbiage” and “B-” respectively.

After 9 iterations

By R. Miller
I had intended to hit the nail, but I’m not a very good aim it seems, and I ended up hitting my thumb. This is a common occurrence, I know, but it doesn’t make me feel any less ridiculous having done it myself. My new strategy will involve lightly tapping the nail while holding it, then it is knocked into the wood enough that the wood itself is holding it straight, and then I’ll remove my hand and pound carefully away. We’ll see how this goes.

Another example: blurry text

After 8 iterations

By R. Miller
Eg, VizWiz: Helping the Blind See

Eg, VizWiz: Helping the Blind See (cont)

By R. Miller
Eg, VizWiz: Helping the Blind See (cont)

Do you see picnic tables across the parking lot? (13s) no (46s) no

What temperature is my 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

By R. Miller
Eg, Image Search

- Using crowds via mobile phones

**CrowdSearch: Exploiting Crowds for Accurate Real-time Image Search on Mobile Phones**, Tingxin Yan et al., MobiSys 2010
Eg, Computer Vision

- How similar is the artistic style?

*Human and Machine Detection of Stylistic Similarity in Art. Adriana Kovashka and Matthew Lease. CrowdConf 2010*
Eg, Database Systems

- Crowd-enabled databases
  - Hybrid human/machine databases
  - Building a database engine that can dynamically crowdsource certain operations

CrowdDB: answering queries with crowdsourcing, Michael J. Franklin et al, SIGMOD 2011
Eg, Crowd + Entity Resolution

- Main idea: Generate candidate pairs using the existing similarity measure.
- Evaluate HITs for candidate pairs.

<table>
<thead>
<tr>
<th>ID</th>
<th>Product Name</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>iPad Two 16GB WiFi White</td>
<td>$490</td>
</tr>
<tr>
<td>r2</td>
<td>iPad 2nd generation 16GB WiFi White</td>
<td>$469</td>
</tr>
<tr>
<td>r3</td>
<td>iPhone 4th generation White 16GB</td>
<td>$545</td>
</tr>
<tr>
<td>r4</td>
<td>Apple iPhone 4 16GB White</td>
<td>$520</td>
</tr>
<tr>
<td>r5</td>
<td>Apple iPhone 3rd generation Black 16GB</td>
<td>$375</td>
</tr>
<tr>
<td>r6</td>
<td>iPhone 4 32GB White</td>
<td>$599</td>
</tr>
<tr>
<td>r7</td>
<td>Apple iPad2 16GB WiFi White</td>
<td>$499</td>
</tr>
<tr>
<td>r8</td>
<td>Apple iPod shuffle 2GB Blue</td>
<td>$49</td>
</tr>
<tr>
<td>r9</td>
<td>Apple iPod shuffle USB Cable</td>
<td>$19</td>
</tr>
</tbody>
</table>

“CrowdER: Crowdsourcing Entity Resolution,” Jiannan Wang, Tim Kraska, Michael J. Franklin, and Jianhua Feng (VLDB 2012)
Eg, Crowd + Entity Resolution

- Possible Human Intelligence Tasks (HITs)

**Pair-based HIT with two records**

**Cluster-based HIT with four records**
Reference

● The Wisdom of Crowds, James Surowiecki, 2004
● Crowdsourcing, Jeff Howe, 2008
● Here Comes Everybody, Clay Shirky, 2008
● Crowdsourcing, Daren Brabham, 2013
● Crowd Computing and Human Computation Algorithms, Rob Miller, 2013
● Crowdsourcing: Challenges and Opportunities, Guoliang Li, 2012