Crowdsourcing

Andrew Kuznetsov
Liu, Yishuo

20170303
Contents

- History
- Applications
- What is Crowdsourcing
- Procedure of MTurk
- Consideration & Challenges
- Three Case Studies
- Future Work in HCI
- Takeaway
Significant Events

1906 Francis Galton
2004 Luis von Ahn - The ESP Game[1]
2005 MTurk Launch
2006 Wired Article [2]
2007 Crowdflower Founding

2. https://www.wired.com/2006/06/crowds/
Applications

Commercial Applications: CAPTCHA / reCAPTCHA[1], DuoLingo
Crowdsourcing knowledge: Wikipedia, Tiramisu, Waze
Micro-task platforms: Mechanical Turk, Daemo, MicroWorkers
Higher-level crowdsourcing: CrowdFlower, TurkPrime, Pybossa
Professional Labor marketplaces: 99Design, Upwork
Citizen Science: Fold it, NASA Clickworkers, LabintheWild
Game with purpose: ESP game
Prediction Markets: PredictIt
Crowdfunding: KickStarter, Indiegogo

What is Crowdsourcing

Definition

"Simply defined, 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. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers."

Concept of Requester and Worker.

“General-purposed” crowdsourcing platform

Three main benefits:

Easy access to scalable, on-demand, inexpensive labour
What is Crowdsourcing (cont’)

Leverage Crowds

• Online surveys
• Recruiting for online/offline experiments
• Dataset for machine learning
• Analyzing text/images
• Gathering subjective judgments

...
What is Crowdsourcing (cont’)

Who Are They? [1, 2, 3]

What is Crowdsourcing (cont’)

What Are the Tasks?

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Rewards</th>
<th>Keyword</th>
<th>#HITGroups</th>
<th>Keyword</th>
<th>#HITs</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>$192,513</td>
<td>castingwords</td>
<td>48,982</td>
<td>product</td>
<td>4,665,449</td>
</tr>
<tr>
<td>collection</td>
<td>$154,680</td>
<td>cw</td>
<td>48,981</td>
<td>data</td>
<td>3,559,495</td>
</tr>
<tr>
<td>easy</td>
<td>$93,293</td>
<td>podcast</td>
<td>47,251</td>
<td>categorization</td>
<td>3,203,470</td>
</tr>
<tr>
<td>writing</td>
<td>$91,930</td>
<td>transcribe</td>
<td>40,697</td>
<td>shopping</td>
<td>3,086,966</td>
</tr>
<tr>
<td>transcribe</td>
<td>$81,416</td>
<td>english</td>
<td>34,532</td>
<td>merchandise</td>
<td>2,825,926</td>
</tr>
<tr>
<td>english</td>
<td>$78,344</td>
<td>mp</td>
<td>33,649</td>
<td>collection</td>
<td>2,599,915</td>
</tr>
<tr>
<td>quick</td>
<td>$75,755</td>
<td>writing</td>
<td>29,229</td>
<td>easy</td>
<td>2,255,757</td>
</tr>
<tr>
<td>product</td>
<td>$66,726</td>
<td>question</td>
<td>21,274</td>
<td>categorize</td>
<td>2,047,071</td>
</tr>
<tr>
<td>cw</td>
<td>$66,486</td>
<td>answer</td>
<td>20,315</td>
<td>quick</td>
<td>1,852,027</td>
</tr>
<tr>
<td>castingwords</td>
<td>$66,111</td>
<td>opinion</td>
<td>15,407</td>
<td>website</td>
<td>1,762,722</td>
</tr>
<tr>
<td>podcast</td>
<td>$64,418</td>
<td>short</td>
<td>15,283</td>
<td>category</td>
<td>1,683,644</td>
</tr>
<tr>
<td>mp</td>
<td>$64,162</td>
<td>advice</td>
<td>14,198</td>
<td>image</td>
<td>1,588,586</td>
</tr>
<tr>
<td>website</td>
<td>$60,527</td>
<td>easy</td>
<td>11,420</td>
<td>search</td>
<td>1,456,029</td>
</tr>
<tr>
<td>search</td>
<td>$57,578</td>
<td>article</td>
<td>10,909</td>
<td>fast</td>
<td>1,372,469</td>
</tr>
<tr>
<td>image</td>
<td>$55,013</td>
<td>edit</td>
<td>9,451</td>
<td>shopzilla</td>
<td>1,281,459</td>
</tr>
</tbody>
</table>

Procedure for MTurk

Work

Qualifications

MTurk Qualifications (Percentage of HITS completed etc...), Custom "Screeners"
Procedure for MTurk (cont’)

Build Tasks

Publish Tasks

Screen Result

Pay

MTurk UI or External Server

Unique Code
Requester Perspective (Link)

Categorize images

Instructions

Consider the image below, then and read the suggested categories for it. For each suggested category, say whether you think the image fits or doesn’t fit.

Task

Results

Created: February 20, 2017
Time Elapsed: 7 days

Assignments Completed: 1,555 / 1,555
Estimated Completion Time: COMPLETE

Batch Progress: 100% submitted
100% published
Workers Perspective (Link)

Transcribe Short Audio Clips
Requester: GoldenAgeTranscription
HIT Expiration Date: Mar 11, 2017 (1 week 3 days)
Reward: $0.08
Time Allotted: 60 minutes

Instructions:

Listen to the short audio clips and transcribe what was said.

- **Verbatim transcription**: Type exactly as spoken. No acronyms or short forms. Example: type "MIT" (instead of "MIT" or "M-I-T")
- **Use English numbers**: Do NOT use any arabic numbers. Example: type "three oh four" (instead of "304") Type "twenty fifteen" (instead of "2015")
- **In-Audible**: Click "In-Audible" if you cannot understand an audio after best effort (noisy, accents or fast/faint speech, non-English, etc.)
- **And... Thank You**

There are 3 audio entries in this task. Please transcribe ALL entries in this task.

Tab: Next Audio; F7: Rewind 3s; F8: Play/Pause;

(type transcription here)

1 - In-Audible (For: very noisy, strong accent, fast/faint speech, none-English)

(type transcription here)

Accept HIT
Skip HIT
Consideration & Challenges

When is it appropriate to use?

- Domain expertise
- Large sample with simple task

Can != Should

- Lab
- Crowdsourcing

Payment

- Quality, Creativity, Demographic
Consideration & Challenges (cont’)

Quality Control

- Worker statistics
- Instruction language
- Verification questions
- Iterate
- Bayesian Truth Serum
- Mechanism Design
Consideration & Challenges (cont’)

Bayesian Truth Serum

\[ P( E | X ) \propto P( E ) P ( X | E ) \]

If people truly hold a particular belief, they are more likely to think that others share the same.

Consideration & Challenges (cont’)

Mechanism Design

● Social transparency [1]
● Social structure [1]
● Skip option[2]
● Rewards[1,2]

Case Study: Assess Wikipedia Quality

Use the crowd to assess Wikipedia articles’ quality

Exp1: rate directly

Exp2: rate with verification questions

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses</td>
<td>210</td>
<td>277</td>
</tr>
<tr>
<td>Invalid</td>
<td>49%</td>
<td>3%</td>
</tr>
<tr>
<td>&lt;1 min</td>
<td>31%</td>
<td>7%</td>
</tr>
<tr>
<td>Median time</td>
<td>1’30”</td>
<td>4’06”</td>
</tr>
</tbody>
</table>

Case Study: Shepherd Turkers

● How to get unskilled crowds to produce better results on complex work?

Task: Write product review

Conditions:

None  Self-assessment  External-assessment

Conclusion: Any sort of assessment help

Case Study: Recruitment

- Personal decisions concerning privacy

Traditional:

2 weeks, 100-200 participants

Students & stuff

MTurk:

2 days, 350 responses, $0.25 per participants

95% white-collar workers
Future Work in HCI

Use Crowdsourcing for HCI

- Recruiting

HCI Research on Crowdsourcing

- Crowd-powered software (Soylent, facade)
- Real-time crowdsourcing (Twitch Crowdsourcing)
- Complex task (Zensors, PlateMate, Flash Teams)
- Crowd / Crowd-team management (Huddler)
- Micro-task interfaces (Compare Me Maybe)
- Improving worker experience (Dynamo, Turkopticon)
Reference


Takeaway

- Growing field
- Broad spectrum of forms and applications
- Great tool for some task with massive, flexible workforce
- Many limitations
- Many open questions