Quality Control Mechanisms in Crowdsourcing Systems

Miao Zhengyuan
Outline

• Motivation

• An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets

• Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk
Motivation

• **Crowdsourcing is getting more popular, thus**
  – how to make crowd more interested in working on Crowdsourcing markets.
  – how to protect the workers’ rights from unfair treatments by employers (e.g. reviews in communities on some bad employers by name blurred workers).

• **Aim to improve the circumstances of the Crowdsourcing market:**
  – Try to make the markets more transparent (e.g. rating for the employers visible).
  – Creating a partial worker-alliance to gain the legal rights for workers.
Outline

• Motivation
• An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets
  • Crowdsourcing and Mechanical Turk
  • Experimental Task
  • Discussion & Conclusion
• Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk
An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets

• **Published**
  • on Jul. 5th, 2011
  • by Jakob Rogstadius, Vassilis Kostakos, Boris Smus (all of these three from Madeira Interactive Technologies Institute University of Madeira), Aniket Kittur (Carnegie Mellon University), Jim Laredo, Maja Vukovic (both from IBM T.J. Watson Research Center)
  • in International AAAI Conference on Web and Social Media Fifth International AAAI Conference on Weblogs and Social Media
Crowdsourcing and Mechanical Turk

- **What is Crowdsourcing? (1 / 2)**
  - was coined in 2005 by Jeff Howe and Mark Robinson.
  - is a sourcing model in which individuals or organizations obtain goods and services (from Wikipedia).
  - is a powerful approach to handle problems that by nature are difficult to solve computationally.
  - different from “Outsourcing” with: diverse specific-levels and diverse work-groups.
Crowdsourcing and Mechanical Turk

- **What is Crowdsourcing?** (2 / 2)
  - There is a similar concept for Crowdsourcing – “witkey”, consists of two words “wit” and “key”.
    - It was coined in 2005 by Liu Feng.
    - Witkey-mode: Transforming people's knowledges, wisdom, experiences and skills into actual income through the Internet.
  - Witkey-mode website: [http://www.epwk.com](http://www.epwk.com) (一品威客)
Crowdsourcing and Mechanical Turk

• **What is Mechanical Turk (MTurk)?**
  
  – Also called as Amazon Mechanical Turk (AMT).
  
  – is a crowdsourcing Internet marketplace:
    – enabling individuals and businesses to coordinate the use of human intelligence to perform tasks.
  
  – is one of the sites of Amazon Web Services and is owned by Amazon (from Wikipedia).
Crowdsourcing and Mechanical Turk

- The Homepage of MTurk and address: https://www.mturk.com

Amazon Mechanical Turk (MTurk) operates a marketplace for work that requires human intelligence. The MTurk web service enables companies to programatically access this marketplace and a diverse, on-demand workforce. Developers can leverage this service to build human intelligence directly into their applications.

While computing technology continues to improve, there are still many things that human beings can do much more effectively than computers, such as identifying objects in a photo or video, performing data de-duplication, transcribing audio recordings or researching data details. Traditionally, tasks like this have been accomplished by hiring a large temporary workforce (which is time consuming, expensive and difficult to scale) or have gone undone.

MTurk aims to make accessing human intelligence simple, scalable, and cost-effective. Businesses or developers needing tasks done (called Human Intelligence Tasks or “HITs”) can use the robust MTurk API to access thousands of high quality, global, on-demand Workers—and then programatically integrate the results of that work directly into their business processes and systems. MTurk enables developers and businesses to achieve their goals more quickly and at a lower cost than was previously possible.

Image Source: https://www.mturk.com/
Crowdsourcing and Mechanical Turk

- Start-up Site

Image Source: https://www.mturk.com/get-started
Crowdsourcing and Mechanical Turk

• How it works on this platform?
  
  – The employers (i.e. the Requesters) put tasks (e.g. labeling objects in an image, transcribing audios) on MTurk.
  
  – The employees (i.e. Workers/Turkers) browse the existed tasks, then take & work.
  
  – Turkers submit the results back to the Requesters.
  
  – Requesters check the results and decide, whether to pay for it. (this problem will be talked in second part.)
Experimental Task

• Main purpose of this paper:
  – To assess the intrinsic and extrinsic motivation on task performance in Crowdsourcing markets.

• Some related works for the same purpose:
  – *Advances in Economic Theory and Econometrics (Gibbons, R. 1997)*: increase extrinsic motivation can elicit higher quality work (e.g. raise payment).
  – *Performance, Pay and Productivity (Lazear, E. 2000)*: workers are more productive when they switched from being paid by time to being paid by piece.
Experimental Task

- The Task to solve: to identify blood cells infected with malaria parasites from the given images.

Three types of the bloods in the image

Example image in the experiment
Sectionalization for the experiment (two “cover stories”)

- One group: non-profit (i.e. charity), “cover story” by the organization *The Global Health Council*.
- Another group: for-profit, “cover story” by the organization *Rimek International*.
- The two “cover stories” have same information except the different organizations.
- For each story three levels of rewards: 0, 3, and 10-cents USD. (each assignment / a piece task earns 0, 3 or 10-cents)
Steps for finishing each assignment:

1) Count the number of malaria parasites in ring-form, having double chromatin dots.

2) Count the total number of blood cells in the image.

- If there was an overlap of cells or the image was ambiguous, then participant should guess or estimate the types of blood.
Metrics and demographics in experiment: (1/3)

• Collected information of each assignment:
  – reported cell count,
  – reported parasite count,
  – time spent,
  – participant ID.

• Personal information by the questionnaire:
  – demographics (e.g. gender, age, education),
  – time registered on MTurk,
  – weekly time spent on MTurk,
  – previous experience with blood analysis,
  – etc.
Metrics and demographics in experiment: (2/3)

- Some demographics results:
  - 158 showed up, i.e. completed at least one assignment (totally 843 people completed the questionnaire).
  - 42% reported lived only in South Asia (including China and India).
  - 35% only in North America (excluding Mexico).
  - Participants from South Asia on average had lower yearly income, higher education and were younger. (compared to those from North America)
Metrics and demographics in experiment: (3/3)

- **The metric formulas:**
  - accuracy metric
    \[
    accuracy = 1 - \frac{1}{2} \left( \frac{|p_{est} - p_{real}|}{p_{real}} + \frac{|c_{est} - c_{real}|}{c_{real}} \right),
    \]
    where \(p\) is parasites and \(c\) is cells.

  - complexity metric
    \[
    complexity = c_{real} + 3p_{real}.
    \]
Hypotheses before experiment:

- **H1**: Tasks in the non-profit (i.e. charity) conditions will be completed faster than tasks in the for-profit conditions.

- **H2**: Tasks in the non-profit (i.e. charity) conditions will be completed more accurately than tasks in the for-profit conditions.

- **H3**: Tasks in high-pay conditions will be completed faster than tasks in low-pay conditions.

- **H4**: Tasks in high-pay conditions will be completed more accurately than tasks in low-pay conditions.
Work Efforts (1/2)

→ Under condition 0-cents were the assignments both not completed.

→ Under condition 10-cents were the works fastest completed compared to other conditions.

*Figure 1: Time taken to complete each condition's batch of assignments*
Work Efforts (2/2)

→ Participants' region of residence (location) affected performance. (yearly income: 5,000$ or below in SA vs 20,000-60,000$ in NA)
→ Payment had also a significant main effect.
Accuracy of tasks

→ Accuracy by non-profit was always better than by for-profit.
→ Accuracy had a generally decreased trend with the increase of sequence no.

**Figure 4:** Mean assignment accuracy by asgmt. sequence number

**Figure 5:** The effect of variations in task complexity on task accuracy (left) and time spent per task (right)
Discussion & Conclusion

• Analysis and discussion from the experiment

  ➔ Intrinsic motivation had a strong positive effect on worker’s accuracy. (When extrinsic factors became main motivator, it didn’t work anymore) —— proposed by prior theory as explanation

  ➔ H2: “Tasks in the non-profit conditions will be completed more accurately...” and H3: “Tasks in high-pay conditions will be completed faster...” were correct.

  ➔ There was sample bias (i.e. lack of diverse sample size) by the experiment.
Discussion & Conclusion

• Conclusion
  → paying people more didn’t lead to increases in their output accuracy.
  → Intrinsic motivation had positive effects on worker’s zeal.
  → The intrinsic motivation frame didn’t impact uptake-speed.
  → When the extrinsic motivation (e.g. reward) was not adequate, performance was likely getting worse.
Any Questions about this paper?
Outline

• Motivation
• An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets
• Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk
  • Turkopticon
  • Introduction
  • Main idea and methods
  • Discussion and conclusion
Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk

• Published
  • between April 27 - May 02, 2013
  • by Lilly C. Irani (UC Irvine, Department of Informatics, CA, USA) and M. Six Silberman (Bureau of Economic Interpretation, Berkeley, CA, USA)
  • in CHI'13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Pages 611-620)
Turkopticon

- Turkopticon is:
  - a third-party platform in Amazon Mechanical Turk (AMT).
  - an extension for Firefox and Chrome (a JavaScript userscript).
  - designed to protect the workers legal rights on Crowdsourcing markets.
  - deployed over four years (now Turkopticon 2 Beta already online).
Turkopticon-Homepage

- If someone wants to see reviews, he must register on Turkopticon first.

- Unluckily, Turkopticon has no more support for Firefox and Chrome extensions. (because this paper wrote with “Turkopticon is a browser extension for Firefox and Chrome”).

Image source: https://turkopticon.ucsd.edu/
Turkopticon 2 Beta-Homepage

- **Which differences from Turkopticon:**
  - Reviews on HITs, not on employers.
  - Flags anonymous.
  - Report objective experiences, not ratings.

Image-source: https://turkopticon.info/
Introduction

• This paper drew on a four-years participant-observation as design activists within AMT worker and technologist communities.

• There were also many experiences and researches in other places (e.g. in HCl, in industry) for enhancing the algorithms of APIs in AMT.
  – For example, Twitter developed a toolkit for running human judgment experiments on AMT.
  – CrowdFlower (another Crowdsourcing platform) built AMT’s APIs and offered crowdsourced data processing tools.
Introduction

• Cause of Internet development Crowdsourcing is more famous worldwide, the work form is not only as software-as-a-service, but also human-as-a-service, even Labor-as-a-service (LaaS).

• Lack of adequate laws for Crowdsourcing that time, so Turkopticon was emerged for gaining workers’ rational rights back.

• As result AMT becomes an infrastructure, on which crowds of Turkers work. But with the progress of AMT the problems between employers and employees come up often. (e.g. delay of payment, reject to pay rewards, etc.)
Introduction

• The first author (Lilly C. Irani) complemented participant-observations (e.g. attending Crowdsourcing conferences, having interviews with employers), and these data was as support of designing and maintaining for Turkopticon.

• Research began in 2008, prior to the growth of popular online worker forums turkernation.com and mturkforum.com.

• 67 respondents answered the question-survey about what they would desire as a “Workers’ Bill of Rights. Points of agreement were as the basis for the design of Turkopticon.
The view of lists of Tasks in AMT

➔ These Tasks are also named HITs (Human intelligence Tasks).
➔ Turkers browse over the list, if find any HIT he wants, then get into that group and pick the task up.
Main idea and methods

• The troubles which workers met:
  – Work was rejected unfairly or arbitrarily.
  – Delay of the payments.
  – “Approval ratings” of themselves fell down.
  – No enough strength to debate the unfairness.
  – Employers’ and Amazon’s lack of response to their concerns.

• The ways which workers wanted to solve the troubles:
  – A forum to build, so that they can air concerns publicly.
  – to build good long-term work relationships with requesters.
  – Partial workers asked for unionization.
Main idea and methods

• Why is Turkopticon so designed?
  • Instead of building a new system prefer an extension for fewer costs.
  • More flexible to sustain and maintain.
  • Easier to browse the reviews of employers generally.
  • Communicate with other workers possible.
Method - averaged ratings of employers

→ Turkers can rank the rating of requesters, so that other workers can decide whether to do the HITs for this requester.

→ The rating consists of four qualities (each value from 0 to 5).

→ Score of "0" means no data for that attribute.

Example: The Turkopticon browser add-on adds information about requesters
Method - averaged ratings of employers

• The four qualities are:
  o **Communicativity**: How responsive has this requester been to communications or concerns you have raised?
  o **Generosity**: How well has this requester paid for the amount of time their HITs take?
  o **Fairness**: How fair has this requester been in approving or rejecting your work?
  o **Promptness**: How promptly has this requester approved your work and paid?
Method - obfuscating workers’ email addresses

Example for comment by worker

→ To prevent employers from retaliating against the employees who wrote reviews about those employers.

→ Problem: If there are malicious comments on requesters or if requester creates account to write many positive comments on themselves?

   o Solution: the Comment Moderation
Method - Comment Moderation

- Turkopticon selects moderators by their reputations and social practices.
- Moderators can judge the comments, to check whether the comments rational, if not they can hide that comment.
- Moderators have the rights to view inside of the design processes of Turkopticon.
- The Moderators can be changed (e.g. next selecting time, or when their reputations fell down).
Discussion and conclusion

- Turkopticon had to update itself, when the Firefox and Chrome released new versions. (now is Turkopticon 2 Beta online, and don’t know if Turkopticon will be still updated.)
- There was a compromise to AMT-Ecosystem, such as to standardize averaged ratings into quantified buckets (not follow author’s original will), aim to remain and attract users.
- Turkopticon had success in activist technologies.
- It would be a plan to develop new media interventions in Turkopticon.
Thanks for your attention!
Any Questions?