Crowdsourcing Experiments, for Better Language Data

Shichang Wang, Chu-Ren Huang, Yao Yao, Angel Chan
CBS, PolyU
10 March 2014

Outline

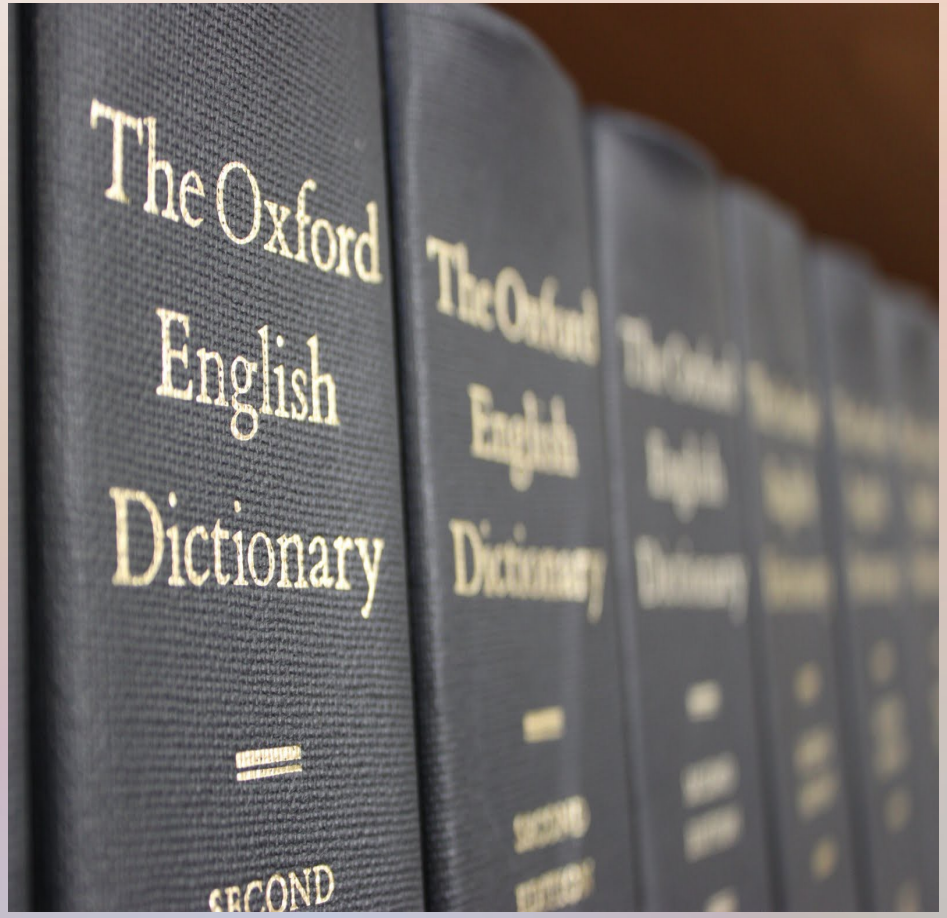
- What is crowdsourcing?
- Crowdsourcing and language data
- An experiment

Section 1: What Is Crowdsourcing?

Definitions

- “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... The crucial prerequisite is the use of the open call format and the large network of potential laborers.” (Jeff Howe, 2006)
- “Crowdsourcing is an online, distributed problem-solving and production model ” (Brabham, 2008)

An “Ancient” Example



- “In the mid-19th century, an open call for volunteers was made for contributions identifying all words in the English language and example quotations exemplifying their usages.”
- “They received over six million submissions over a period of 70 years.”

A Modern Example

- "a multilingual, web-based project to **create a free content dictionary of all words in all languages**. It is available in 158 languages and in Simple English."
- "allows almost anyone with access to the website to create and edit entries."

The Forms of Crowdsourcing

- Games with a Purpose
- Amazon Mechanical Turk (MTurk)
- Wisdom of the Crowds

Section 2: Crowdsourcing and Language Data

Why Crowdsourcing Can Bring Better Language Data?

- Big Sample
- High Diversity
- Low Cost
- Automation
  - Automatic language resource construction
  - Self-evolving language resource

The Effects of Anonymity

- Positive Effects
  - Participants are better protected from privacy disclosure
  - Participants can be more open which enables us to collect sensitive data
- Negative Effects
  - The cost of cheating is low
  - Cannot analyze/control the identities

The Limitations of Crowdsourcing

- The Noise Problem
  - Participants may be not as careful as they are in lab experiments
  - Spammers for “quick cents” (human & robot)
- The Experiment Control Problem
  - The identity/background information of the participants is usually not accessible
  - Difficult to realize special experiment settings
- No Perfect Method, But Only Best-fit Method

Section 3: An Experiment

Objectives

- To check the feasibility of collecting Chinese language data from common crowdsourcing platforms based out of China, e.g., Amazon Mechanical Turk (Mturk) and Crowdflower.
- To identify and solve the issues in study design and data quality for the benefit of future practices of the same sort.

Design and implementation

- Survey-based experiment
- The questionnaire
  - 46 questions
  - 3 parts
    - Part 1: Screening questions
    - Part 2: Chinese word segmentation task
    - Part 3: Semantic similarity judgment task
- *See Questionnaire.htm*

Platform and Payment

- Compared with MTurk, Crowdflower was more efficient and effective in recruiting Chinese-speaking participants.
- Each participant was only allowed to submit the job once.
- US$0.25/response

Valid/invalid Responses, the Criteria

- The screening questions in Part 1 were correctly answered
- The answers in Part 2 followed the correct format
- The completion time was equal or greater than 5 minutes

Process/Overview

- Aimed to collect a sample of 200 responses.
- After 135 responses were collected, we found a serious spammer problem, so we paused the experiment to seek solution. (Stage 1)
- Wrote a monitor program based on the API of Crowdflower which could detect and resist spammers automatically.
- The experiment was resumed after 2 months, with the monitor program; collected another 65 responses. (Stage 2)

Process/Stage 1

- 135 responses were received
  - 88 (65.19%) valid
    - 81 (92.05%) contributed by participants who claimed to be from Mainland China
    - 7 (7.95%) by participants from Hong Kong
  - 47 (34.81%) invalid

Process/Stage 2

- 65 responses were received
  - 54 (83.08%) valid
    - 46 (85.19%) contributed by participants from Mainland China
    - 8 (14.81%) by participants from Hong Kong
  - 11 (11.92%) invalid
- The data obtained from Stage 1 and Stage 2 were highly similar, suggesting that the experiment was replicable.

Data

- Obtained 200 responses, 142 (71%) were valid
- The valid responses showed high consistency
- For example, we analyzed 127 responses contributed by participants from Mainland China
  - Part 2 (Chinese word segmentation): the average consistency is 74.30% (SD=12.94%)
  - Part 3 (Semantic similarity judgment): the average consistency is 58.46% (SD=21.97%)
  - Largely consistent with our expectations

Data/Examples

分词结果	数量	百分比
他/十分/高兴	122	96.06%
他/十分高兴	3	2.36%
他十分/高兴	1	0.79%
他/十/分高兴	1	0.79%
总计	127	100%

Data/Examples

分词结果	数量	百分比
这/只不过/是/个人/问题	89	70.08%
这/只/不过/是/个人/问题	14	11.02%
这/只不过/是/个人问题	11	8.66%
这/只不过是/个人/问题	5	3.94%
这只/不过/是/个人/问题	3	2.36%
这/只/不过/是/个人问题	1	0.79%
这/只/不/过/是/个人/问题	1	0.79%
这只/不过/是个人/问题	1	0.79%
这只不过/是/个人问题	1	0.79%
这/只不过/是个人/问题	1	0.79%
总计	127	100%

Data/Examples

	东西		地步		漂亮		风度		出息		利索	
相似度	东	西	地	步	漂	亮	风	度	出	息	利	索
1	115	121	94	100	79	63	109	84	97	110	80	98
2	2	2	12	11	15	32	8	29	13	7	15	12
3	1	1	8	8	10	15	3	7	4	3	15	6
4	0	1	3	2	9	7	0	2	3	0	3	0
5	8	1	9	4	11	5	7	3	8	3	9	3
?	1	1	1	2	3	5	0	2	2	4	5	8

Data/Examples

	帮助		衣服		告诉		制作		兑换		灾祸	
相似度	帮	助	衣	服	告	诉	制	作	兑	换	灾	祸
1	6	4	2	32	20	19	4	12	3	3	3	2
2	2	13	8	29	23	41	22	25	13	8	5	11
3	8	7	12	19	24	30	20	31	13	16	16	21
4	23	38	27	24	26	21	43	33	44	42	41	43
5	88	63	78	20	32	13	36	24	54	56	62	50
?	0	2	0	3	2	3	2	2	0	2	0	0

Conclusion

- It is feasible to collect Chinese language behavior data from a crowdsourcing platform like Crowdflower which is based out of China.
- When used properly, the crowdsourcing technology may provide an efficient, effective and economic way to collect Chinese language behavioral data with reliable data quality.

Thank you very much!

References

- Brabham, Daren C. "Crowdsourcing as a model for problem solving an introduction and cases." Convergence: the international journal of research into new media technologies 14.1 (2008): 75-90.
- Howe, Jeff (June 2, 2006). "Crowdsourcing: A Definition". Crowdsourcing Blog. Retrieved March 8, 2014.
- Wang, A.; Hoang, C. D. V. & Kan, M.-Y. (2013), 'Perspectives on crowdsourcing annotations for natural language processing', Language Resources and Evaluation 47, 9--31.
- http://en.wikipedia.org/wiki/Wiktionary
- http://en.wikipedia.org/wiki/Crowdsourcing