

# 630 HW6 Project Report

## Questions and Answer Part:

### Part 1: Measuring Inter-Annotator Agreement

1. Compute alpha for just the group members and report it
  - a. 0.299 (fair agreement)
2. Compute alpha for just the crowd workers and report it
  - a. 0.312 (fair agreement)
3. In a 2-3 sentences, comment on the difference (if any) between your group's agreement and the crowd's agreement. Which is higher?
  - a. The crowd's agreement is higher, although we're quite close. This indicates that crowds and the group have similar criteria, that is, our instructions work.
  - b. From the perspective of the thumb of rule, both alphas indicate fair agreement.
4. Compute the correlation between all pairs of annotators in your group and report these in a table. You should also report the average correlation for the group.

	Wen Xin	Linyan Ge	Bicheng Xu
Wen Xin	/	0.4113861223806193	0.34287678110018077
Linyan Ge	0.4113861223806193	/	0.39723522971958747
Bicheng Xu	0.34287678110018077	0.39723522971958747	/

- a. Average correlation: 0.384
5. Compute the average correlation between all pairs of crowd workers. Here, you'll only compute the correlation on the items they *both* annotated.

Average correlation between all pairs of crowd workers: 0.563

We attached our correlated matrix with files we submit.
6. In 2-3 sentences, how does the correlation scores compare with the alpha values? Which seem like better estimates?
  - a. Alpha values is lower than Pearson value.
  - b. Alpha value is better for two reasons. On the one hand, the scores are dichotomous, and alpha value is the most common measure of internal consistency. On the other hand, it is more versatile, which means it can handle three or more answers per variable. In this project, Likert type scale answers is a good example.

## Part 2: Examining the disagreement

1. Include the text from each of these five instances and what the ratings were for each group

post_text	group_mean	crowd_mean	mean_diff	
Slept like ass, just had a vegetable smoothie which similarly tasted like ass, but I need me those vitamins and good stuff, so a necessary evil. Thanks, comrade.	-2.3333333	0.66666667	3	
Unless your boobs are pretty big. Ive got some great bras and struggle with finding bathing suit tops that are supportive and dont tie around the neck and not super expensive :( my bras are sturdy af lol	2	-0.6666667	2.66666667	
Heres the original ballad [NSFW]. Fuck was bowlдерized as f-k.	0	2.33333333	2.33333333	
Thanks, one of the best bikes Ive had yet. Was able to cross by going down the washout finding a path through all the fallen trees and such until I found a spot to get up the bank.	-1.6666667	0.66666667	2.33333333	
Lower your expectations :)	1.66666667	-0.3333333	2	

2. Describe why you think the two were different.
  - a. The group pays more attention to the syntactic and spelling error than the crowd seems.
  - b. Culture difference. The group members are all Chinese, so sometimes cannot get the true meaning of a reply.
  - c. By chance. Or the crowd is not doing good work.
3. In discussion with all group members, adjudicate the rating and decide what the final “true” rating is, examining all the evidence. If you changed your score, in one sentence, state what led you to do this.

post_text	group_mean	crowd_mean	mean_diff	final
Slept like ass, just had a vegetable smoothie which similarly tasted like ass, but I need me those vitamins	-2.33	0.67	3	-2

and good stuff, so a necessary evil. Thanks, comrade.				
Unless your boobs are pretty big. Ive got some great bras and struggle with finding bathing suit tops that are supportive and dont tie around the neck and not super expensive :( my bras are sturdy af lol	2	-0.67	2.67	1
Heres the original ballad [NSFW]. Fuck was bowlderized as f-k.	0	2.33	2.33	-1
Thanks, one of the best bikes Ive had yet. Was able to cross by going down the washout finding a path through all the fallen trees and such until I found a spot to get up the bank.	-1.67	0.67	2.33	1
Lower your expectations :)	1.67	-0.33	2	1

Changes:

1. Change to “-2”, as several ‘ass’
2. 1. It’s related to the context, and is polite, but spelling errors
3. Change to ‘-1’. ‘Fuck’ is not civil.
4. Change to ‘1’. Some spelling errors.
5. 1. We think we are right.

### Part 3: Follow-up improvements to instructions

Based on your analysis of annotator agreement and the disagreements, what improvements could be made to your instructions? In a series of bullet points, describe in 1-2 sentences specific improvements or changes that could be made to increase annotator agreement or correct for gaps and edge cases.

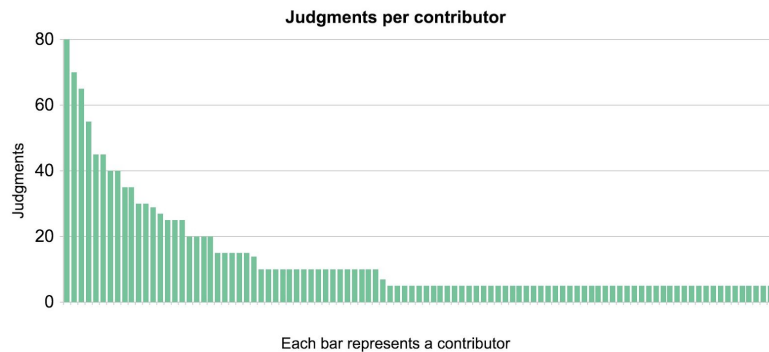
- People should consider the context of the reply1. For example, some context seems indecent, while the reply1 is not as indecent as the context. At this time, reply 1 should be considered as civil
- Some short context which have little information, but use friendly emojis should consider civil

## Details of CrowdSourcing task Part:

### Contributor:

The contributors that we used are all qualified contributors, which are in level 1. In this way, we can collect our data in the fastest way.

The trusts of most judgements of contributors decreases exponentially , which seems not so good. However, there are many great contributors who have both large amounts of judgements and high trust value.

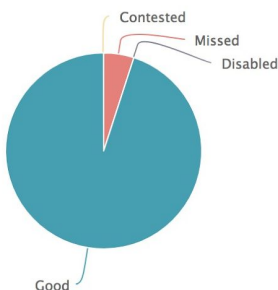


Contributor ID	Judgments ▼	Trust
44,419,709	80	85%
44,122,640	70	83%
22,346,229	65	100%
38,128,837	55	93%
44,407,873	45	85%
42,471,295	45	85%

However, for the contributor satisfaction, the overall score is 3.1 / 5. And the scores for each items are here: 3 / 5 for Instructions Clear, 2.9 / 5 for Test Questions Fair, 2.6 / 5 for ease of Job and 2.6 / 5 for the payment. Except the ease of job, test questions fair the payment, which we cannot improve, it shows that we could improve the instructions when we do crowdsourcing next time.

#### Test Questions ⓘ

Number of test questions: 20



#### Contributor Satisfaction ⓘ

Number of participants: 10

**3.1** / 5

Overall

**3** / 5

Instructions Clear

**2.9** / 5

Test Questions Fair

**2.6** / 5

Ease Of Job

**2.6** / 5

Pay

## Quality Control:

We used 30 mins seconds as our minimum time per page.

### Minimum Time Per Page

#### Minimum time it should take a contributor to complete a page of work

When triggered, the contributor will be removed from your job.

30

minimum seconds

## Quality:

For the 20 questions that we set before, the missed rate is about 25 percent each, which seems a little bit high. It might because we are Chinese people who are not familiar with English Internet environment.

### Quality

Reports ▾

Create More

ID	% CONTESTED	% MISSED	JUDGMENTS	LAST UPDATED	ENABLED ▾
1672478808	<div><div></div></div>	<div><div></div></div>	32	2 days ago	<input checked="" type="checkbox"/>
1672467065	<div><div></div></div>	<div><div></div></div>	40	2 days ago	<input checked="" type="checkbox"/>
1672475823	<div><div></div></div>	<div><div></div></div>	43	2 days ago	<input checked="" type="checkbox"/>
1672481560	<div><div></div></div>	<div><div></div></div>	47	2 days ago	<input checked="" type="checkbox"/>
1672478169	<div><div></div></div>	<div><div></div></div>	42	2 days ago	<input checked="" type="checkbox"/>
1672470826	<div><div></div></div>	<div><div></div></div>	36	2 days ago	<input checked="" type="checkbox"/>
1672473962	<div><div></div></div>	<div><div></div></div>	43	2 days ago	<input checked="" type="checkbox"/>
1672476475	<div><div></div></div>	<div><div></div></div>	44	2 days ago	<input checked="" type="checkbox"/>
1672465137	<div><div></div></div>	<div><div></div></div>	39	2 days ago	<input checked="" type="checkbox"/>
1672468772	<div><div></div></div>	<div><div></div></div>	45	2 days ago	<input checked="" type="checkbox"/>
1672480241	<div><div></div></div>	<div><div></div></div>	38	2 days ago	<input checked="" type="checkbox"/>
1672479767	<div><div></div></div>	<div><div></div></div>	43	2 days ago	<input checked="" type="checkbox"/>
1672472091	<div><div></div></div>	<div><div></div></div>	36	2 days ago	<input checked="" type="checkbox"/>
1672464387	<div><div></div></div>	<div><div></div></div>	39	2 days ago	<input checked="" type="checkbox"/>
1672474269	<div><div></div></div>	<div><div></div></div>	42	2 days ago	<input checked="" type="checkbox"/>
1672474556	<div><div></div></div>	<div><div></div></div>	44	2 days ago	<input checked="" type="checkbox"/>
1672469320	<div><div></div></div>	<div><div></div></div>	42	2 days ago	<input checked="" type="checkbox"/>
1672472691	<div><div></div></div>	<div><div></div></div>	34	2 days ago	<input checked="" type="checkbox"/>
1672473576	<div><div></div></div>	<div><div></div></div>	39	2 days ago	<input checked="" type="checkbox"/>
1672465843	<div><div></div></div>	<div><div></div></div>	43	2 days ago	<input checked="" type="checkbox"/>

Payment:

The number of rows displayed on each page is 5 and the payment amount per page is 10 cents. The overall cost is 20 dollars.

Payment Settings

Set the number of rows displayed on each page and the payment amount per page.

Rows per Page ⓘ

5

Price per Page ⓘ

10

 ¢

Monitor

