

COLX 523 Week 3

Project manager of the week: Andrew

Tasks (points allocated)

1. Annotation + explanation + code (9) [Team]

- ~~intermediary files produced during the annotation process [Toshiko]~~
- ~~annotation [Andrew, Jiang, Yuesheng]~~
- ~~the final version of the annotation [Jiang]~~
- ~~code used to prepare the data for annotation [Toshiko]~~
- ~~code used to convert the raw annotation into the final annotation [Jiang]~~
- brief discussion of how the annotation process went [Andrew]

2. ~~Interannotator agreement study (6) [Jiang]~~

~~Note if your agreement is low and it appears that it is low because you didn't take steps to ensure quality, you will lose points here even if the results of the study are accurate.~~

- ~~Use the 'keyword_aspect' extracted from the sentence keywords as a optional reference.~~
- ~~Apply Cohen's methods~~
- ~~Three groups of annotations: yuesheng/Andrew/Jiang~~

3. Experimenting with annotation options (Optional) (2) [Andrew]

Logs of time required for each annotator and discuss

4. Plan for the interface (5) [Yuesheng]

How can we prevent us from unreliable annotations?

- Shuffle the sentences → ~~group sents by the title~~
- ~~Add a column to provide background info for annotators (eg. Green New Deal (GND) proposals call for public policy to address climate change.. → political efforts → this is about politics)~~
- ~~Post a small sample of sents first, collect annotations from several Turkers, pick the reliable Turkers for the rest of the sentences~~
- ~~Set requirements for annotators, such as university grads or higher, age limit?~~
 - → We will do annotations
 - 500 annotations each by Andrew & Yuesheng
 - 250 annotations each by Jiang & Toshiko
 - Would we be able to split up annotations this way and include Toshiko and Jiang in the agreement study -- as separate annotators? --Yes, we would have two separate studies, one Andrew + Yuesheng + Jiang, one Andrew + Yuesheng + Toshiko. For the optional question -- how much time did we spend discussing vs. annotating?
 - **Log the amount of time we spend on annotating.**
 - What can we do for the optional question given that we are not crowdsourcing?
 - We can discuss a few ambiguous examples
 - Can ask for extension Friday at noon if we need 24-hour extension for milestone 3; for milestone 4, our deadline is on Monday, but can ask for a couple extra days if needed.

-Shuffle the sentences -- Toshiko

-Annotation + explanation + code -- Andrew and Toshiko

-Interannotator agreement study -- Jiang

-Plan for interface -- Yuesheng

-To do later: decide on task distribution for week 4

Deadlines:

- Tuesday at 2:00: discuss first ten annotations (see random numbers on slack)
- Toshiko, Tuesday: shuffle the rows

-Wednesday night: finish annotations