

COLX 523 Week 2

Project manager of the week: Qichao (Jiang)

Tasks (points allocated)

1. Corpus collection code (6) [Toshiko, Jiang]
2. Corpus collection [Toshiko]
3. ~~Corpus analysis (Optional; 2)~~
4. Annotation plan (5)
 - Reasoning [Andrew, Toshiko, Jiang]
 - Mechanical Turk [Yuesheng]
 - Writing [Jiang, Yuesheng]
5. Annotation materials (4)
 - Guideline [Andrew], data preparation [Andrew, Toshiko], Mechanical Turk [Yuesheng]
6. Revised proposal (7) [Andrew]
7. Prompt completion (4) - open an issue in the repo [Jiang]

~~Web scraper specifications~~

~~Corpus data structure:~~

~~I chose JSON because it can handle a list/dict/set of variable lengths as one feature/cell
→ good for a discussion thread that has a tree-like structure (directed graph) of a
variable number of nodes (i.e., comment).~~

~~Examples (a thread of 3 vs. 5):~~

~~<https://github.com/chrdey/change-my-view-modes/blob/master/v2.0/positive/0.metadata>~~

~~<https://github.com/chrdey/change-my-view-modes/blob/master/v2.0/positive/10.metadata>~~

~~[a](#)~~

File organization

For Submission data: submission_dict.json

Reddit/paw	Column name	Note
id	submission_id	

author.name	submission_author	
url	submission_url	
title	submission_title	
selftext	submission_body	
upvotes		
comments		

For comment thread: delta_thread_dict.json

```

▼ root:
  ▶ 9syrun-e8snnp: [] 2 items
  ▶ axbnv4-ehsmb3: [] 6 items
  ▶ b3t23h-ej2740f: [] 2 items
  ▶ dy6sbf-f7z7e4o: [] 6 items
  ▶ e976ru-fagy2fu: [] 4 items

```

key: "submission_id-comment_id" (the identifier for a thread)

value: a list of comments contained in that thread (tree/directed graph structure)

In the above example, there are 5 threads. The first thread has 2 comments and the second has 6.

Annotation plan (TBD)

My suggestion is to follow the annotation plans proposed by these two articles:

Hidey et. al., "Analyzing the Semantic Types of Claims and Premises in an Online Persuasive Forum" <https://aclanthology.org/W17-5102/>

Chakrabarty et. al., "AMPERSAND: Argument Mining for PERSuAsive oNline Discussions" <https://arxiv.org/abs/2004.14677>

Specifically, in the two-step order (details need to be discussed - Skip "logos/pathos/ethos?")

Step 1: Argumentation tags: "Claims", "Premises" by following this:

<https://github.com/chridey/change-my-view-modes/tree/master/original>

Step 2: Relation tags: "Support", "Agreement", "Attach", "Rebuttal", "Undercutter", "Partial attack"

https://github.com/chridey/change-my-view-modes/blob/master/v2.0/argumentative_relation_annotation_guidelines.docx

Sample annotations

<https://github.com/chridey/change-my-view-modes/tree/master/v2.0> provide the entire corpus

Annotation proposed so far :

(Vote For the annotation plan you like, you can vote for more than one plan. If you didn't vote for a plan, please comment why you don't like it.)

1. Argument annotation:

Annotate each comment from the 5 classes below. Since we are at the "Change My View" subreddit, most comments are rebutting the OP's opinion (original post). The first three classes should be summative enough about how they rebut the OP. And we include the last two classes to deal with outliers and ambiguity. We can filter out some long comments to simplify the work since Toshiko has collected around 2M words in the subreddit so far:

- rebut a conclusion
- rebut a premise
- undercut an argument
- support the original post (which is rare)
- other

_____ (Voter: Jiang)

_____ (Comment:)

2. Sentiment Annotation Based on Aspects

Data: scraping

- Aspects:
- Polarity: positive/negative/neutral

3. Sarcasm binary annotation