

Collective Annotation of Linguistic Corpora

Ulle Endriss

Institute for Logic, Language and Computation

University of Amsterdam

[joint work with Raquel Fernández, Justin Kruger and Ciyang Qing]

Computational Social Choice

I work on *computational social choice*, which is concerned with the design and analysis of methods for collective decision making:

- voting and preference aggregation
- fair allocation of resources
- judgment aggregation

Techniques used include logical and probabilistic modelling, game-theoretical analysis, algorithm design, complexity analysis, philosophical scrutiny, and data-driven studies.

Collecting Raw Annotations: Crowdsourcing

amazonmechanical turk Artificial Intelligence

Your Account | **HITs** | Qualifications | **244,501 HITs** available now

Sign In

All HITs | HITs Available To You | HITs Assigned To You

Find HITs containing [] that pay at least \$ 0.00

for which you are qualified
 require Master Qualification GO

Timer: 00:00:00 of 15 minutes

Want to work on this HIT? **Accept HIT** | Want to see other HITs? **Skip HIT**

Total Earned: Unavailable
Total HITs Submitted: 0

1. Yes-No Questions [Show examples](#)

Questions that *have the standard form* of a question and that *could be* answered by saying "yes" or "no" (Careful! They are *not always answered in this way*. It only matters whether they could).

2. Wh Questions [Show examples](#)

Questions that *have the standard form* of a question and that ask for *specific* information by means of a question word such as "what", "who", "which", "when", "where" or "how".

3. Declarative Questions [Show examples](#)

Questions that *don't have the standard form* of a question (they look more like statements) but *nevertheless ask for some answer*, which could be a "yes"/"no" answer or more specific information.

4. Rhetorical Questions [Show examples](#)

Questions that *do not need to be answered*. They can *have the form of any of the question types above*, but they are asked only to *make a point* (often negative), for the sake of encouraging the listener to consider an issue.

In this task you are asked to classify the questions in 10 fragments of dialogues, according to the definitions on the left (with examples):

Read the definitions of different types of questions on the left carefully, as well as the examples that follow. Please choose the type that is closest to the usage of the question marked in bold in each dialogue fragment below. (You should always classify what is marked in bold, even if sometimes it is without a question mark!)

Dialogue 1.
 A: and the other one doesn't.
 A: And you're right, they do get bored, uh, really fast, if they already know what you're talking about.
 A: What do you propose that they do?
A: What, what is your suggestions?
 B: The educators need to be a little bit more open minded as well as innovative in dealing with, uh, the various students to get the maximum potential out of the person.
 A: Uh-huh.
 A: Out of each child.

Yes-No Wh Declarative Rhetorical

Want to work on this HIT? **Accept HIT** | Want to see other HITs? **Skip HIT**

Aggregating Raw Annotations

Aggregating information from individuals is what *social choice theory* is all about. Classical case: aggregation of preferences in an election.

F : vector of individual preferences \mapsto election winner

F : vector of individual annotations \mapsto collective annotation

Three Ideas for Sophisticated Aggregation Rules

Naïve approach: *majority voting*. We have developed three more sophisticated (families of) aggregation rules:

- (1) *Bias-Correcting Rules*: Weighted majority rule, giving annotators *lower weight* on categories they use with *higher frequency*.
- (2) *Agreement-Based Rules*: Estimate each annotator's *accuracy* by comparing with the majority opinion. Then use weighted majority, with *weights* calibrated by maximum likelihood estimation.
- (3) *Greedy Consensus Rules*: Go through items by *majority strength*, always following the majority opinion, but progressively *eliminate* annotators who disagree with the majority too often.

Last Slide

Our results show that sophisticated aggregation rules can increase quality significantly (or: maintain quality with fewer annotators).

Our papers and crowdsourced data are available here:

<http://www.illc.uva.nl/Resources/CollectiveAnnotation/>

U. Endriss and R. Fernández. Collective Annotation of Linguistic Resources: Basic Principles and a Formal Model. Proc. ACL-2013.

J. Kruger, U. Endriss, R. Fernández, and C. Qing. Axiomatic Analysis of Aggregation Methods for Collective Annotation. Proc. AAMAS-2014.

C. Qing, U. Endriss, R. Fernández, and J. Kruger. Empirical Analysis of Aggregation Methods for Collective Annotation. Proc. COLING-2014.