Assignment 1

Emotion Analysis 22/23

Publication: 2022-11-08
Submission Deadline: 2022-11-20
Live Discussion Session: 2022-11-22

Roman Klinger

- **Groups:** Working in groups of up to three people is encouraged, up to four people is allowed. More people are not allowed. Copying results from one group to another (or from elsewhere) is not allowed. Changing groups during the term is allowed.
- **Grading:** The grade for this class is a combination of the assignment submissions and a written exam at the end of the term. For the assignments, we will consider the following aspects to assign a grade: (1) has the data source/domain source be explained and justified in the presentation, (2) is the procedure well motivated and explained, (3) are the slides understandable and easy to follow, (4) do the students critical reflect on their work, (5) did they contribute own creative ideas, (6) was the presentation understandable and well organized/understandable (including an oral presentation if presented), (7) the outcome has been analyzed properly.
- **Slides and additional document:** We acknowledge that it might be challenging to prepare a slide presentation that is well-suited for presentation and well-suited to be understood in isolation. The slide submission is mandatory. If you prefer to additionally write a short document explaining additional aspects of your work, feel free to do so.
- **Submission:** First make a group in Ilias, then submit the PDF file(s). Write all group members on the first pages of both PDFs. We might otherwise not be able to associate your submissions with the group members. If you are technically not able to make a group (it seems that happens on Ilias from time to time), do not submit a PDF multiple times by multiple people – only submit it once.
- **Make it understandable:** Do the best you can such that we can understand what you mean. We cannot grade what you did – only what we see what you did.

Goal

The goal of this assignment is to develop an own corpus, annotated for emotions, and discuss and reflect on various design choices that are required.

Material

Please access the annotation spreadsheet at the following link or download it from the website (this is an example for an annotation environment, you can also use something else if you prefer): `https://tinyurl.com/y3w6fmvb`

Step 1: Task Specification

Before getting started, discuss with your team members which data type you would like to work on. Examples are Tweets, SMS, Emails, News Headlines, or Reddit posts. We will not ask you to submit the data to us, so feel free to use data that you do not want to share (as long as you are ok with sharing it inside of your group and showing examples).

Further, decide on the emotion model you plan to follow. Do you annotate discrete emotion names (which set)? Intensities? Confidence values? Valence/arousal/dominance?

These decisions come together – when deciding on a domain/data source, this might influence the choice of the label/variable set. While making these decisions, also look at some example instances.

Step 2: Prepare Annotation Environment

Now you know what you want to annotate and with which concepts. Now, please, prepare your annotation environment with at least 50 instances from the text source you have chosen. You can use the spreadsheet to which we share a link above, but you are also free to use something else. If you use the environment, be sure to make a copy before putting data in (if working online).

Step 3: Annotation Guideline

As a starting point to prepare the annotation guidelines, annotate at least 10 instances from the annotation environment together. Discuss your decisions and how you came to your results.

When you are done, write a small document in which you describe the task in a way that somebody else would be able to annotate for you. That can include technical aspects (click here, click there) and a description of classes/variables to chose from.

Step 4: Annotate Independently

Now we simulate the first independent annotation round. Make a copy of the annotation environment for each group member and annotate at least 40 instances independent of each other. When you are done, calculate inter-annotator agreement scores (for each emotion class, or correlations for each variable, pairwise for all annotators or together for all of them). Also discuss differences in the annotation and try to understand why you disagreed.

The goal of this part of the exercise is: Understand the challenge of emotion annotation and particularly the difficulties that you were faced with.

If you work alone, do the annotation twice and calculate the inter-annotator agreement between the two rounds.

Step 5: Analysis

Prepare corpus statistics, either for an aggregated version of your corpus (majority vote for each instance) or for each annotator. How many instances were annotated with which class? Are there patterns of disagreement?

Do not only rely on numbers, also have a look at the instances and analyze them qualitatively. Are there patterns in agreement/disagreement? Can you find reasons for disagreement? Do the emotions cover all phenomena that you expect from the particular domain you have chosen? Please reflect your work critically.

Step 6: Prepare Submission

Prepare roughly 6 slides for the submission of the exercise, report the data source, selection procedure, motivate and report the task specification. Report corpus statistics and inter-annotator agreement scores, discuss annotation difficulties. Show some examples. This presentation should be for your fellow students, not for the teachers of the course – they all worked on a similar task, but only you have chosen this particular instance. Tell them why you made the decisions you made, let them benefit from your experiences.

Be prepared to present these slides in 5–10 minutes. If you want to report additional things, please prepare an additional document in which you share further information. The submission of the slides is mandatory, the submission of an additional document is optional.

Submit the documentation of your work in PDF format.