

Assignment 4

Emotion Analysis 22/23

<https://www.emotionanalysis.de/>

Publication: 2023-01-10

Submission Deadline: 2023-01-29

Live Discussion Session: 2023-01-31

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 an oral exam at the end of the term. For the assignments, we will consider the following aspects to assign a grade: (1) is your work well-motivated, (2) is there a research question or hypothesis, (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. But 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 do not only grade what you did – we grade what we see what you did.

The goal of this assignment is that you familiarize yourself with the concept of role labeling or stimulus detection. You can choose between two possible overall tasks, either **annotation** (similar to assignment 1) or **modelling** (similar to assignment 2). You only need to work on one of both.

Option 1: Role Labeling/Stimulus Annotation

This option is suitable for you if you are more interested in the corpus creation and analysis process than in the computational modelling.

Step 1: Task Specification and Corpus Choice

Before getting started, discuss with your fellow team members what kind of data you would like to work on (news, headlines, social media, literature, ...). You can also decide to reannotate the data that you worked on in Assignment 1. Further, decide what you want to annotate. That should be the emotion category, and a subset of roles (stimulus, experiencer, emotion cue). The concrete choice and what might be interesting and relevant also depends on the domain data decision. You should at least annotate two span-based concepts. Ideally, come up with a research question that you would like to answer based on your annotation.

Step 2: Prepare Annotation Environment

Now prepare the annotation environment. As a minimal solution, that could be a spreadsheet with columns for the text (that you prepare) and each role (which is filled by the annotators). The role column could then simply be filled with the text of the role, copy-pasted from the original text. That might be seen as an oversimplification, because one role text could occur in several places, and it might be ambiguous (for instance, if it's purely "it" as a stimulus). An alternative would be to use a proper annotation environment, like WebAnno, Incentive.

Step 3: Annotation Guidelines

Write a small text that explains to the annotators how to annotate, as in Assignment 1, the goal would be that somebody else would be able to do the job for you.

Step 4: Annotate independently and Analyze

Annotate at least 30 instances, independently from each other. Calculate inter-annotator agreement (e.g., via token-based Cohen's kappa, or by assuming that one annotator is the "gold" and the other is the "prediction" – then you can calculate an "inter-annotator F score").

Analyze the differences, prepare corpus statistics, and submit a 10 minutes slide presentation and optionally an explaining document (same procedure as before). Please also volunteer for the presentation, particularly if you did not present yet.

Option 2: Develop role-labeler/stimulus detector

This option is suitable for you if you did implement sequence-labeling methods (like named entity recognition) or relation-detection methods before, or you are willing to learn how to do that. Possible starting points, if you do not have prior experience are

- https://www.nltk.org/_modules/nltk/tag/crf.html
- https://huggingface.co/transformers/model_doc/roberta.html
- <https://www.aclweb.org/anthology/P16-1101.pdf>
- https://pytorch.org/tutorials/beginner/nlp/sequence_models_tutorial.html

Step 1: Data Choice

Similar to Assignment 2, we provide you with a corpus file in which those corpora that we are allowed to make available to you are available in a standardized format. This file is associated to the publication at <https://www.aclweb.org/anthology/2020.peoples-1.12.pdf> and accessible via <https://github.com/sarnthil/extract-emotion-data>.

You can either use all of these corpora or just a subset of them. You can also decide to use a different corpus (either one that we discussed in class or one that you found elsewhere). Make sure to explain your decisions. Come up with a question that you would like to answer with your experiments.

Step 2: Method Decision and Implementation

Decide what kind of sequence labeling (or clause classification) method you would like to use. You should have at least two different configurations, for instance a simple version and an advanced version. What the relation between these methods is depends on the research question that you posed in the previous step.

Implement the method and apply it to the corpora that you have chosen. Note that this quickly becomes complicated – as we discussed in class, both evaluation and recognition are clearly non-trivial. Make sure to first implement a simple version that runs and does something, because trying to implement all possible more sophisticated options. It's better to have something simple that works than several complicated attempts, but none of them works from start to the end.

Step 3: Evaluate your method

In the evaluation (which you should obviously do on an independent test set or via cross validation), report span-based evaluation measures. Token-accuracy/precision/recall/F are not appropriate for the task of sequence labeling, in which the relevant instances are typically longer than one token.

Compare your two variants of the sequence labeling method, discuss the results, answer the question that you posed at the beginning.

Submit a 10-minutes presentation and optionally another PDF file (same procedure as in assignment 1 and 2).