ML-based Emotion Role Labelling

Emotion Analysis
Assignment 4

University of Stuttgart
Germany

Felix Bühler
Carlotta Quensel
Max Wegge

research question
data choice

Does the choice of training data influence the result of the classifier and how?

- emotion roles are determined semantically but this information is partially included in the syntactic structure

- label the emotion target which is often an NP (for example the person or institution, the emotion is directed at)

- How will corpora with very different syntax change the trained algorithm:
  - GoodNewsEveryone: news headlines, which are abbreviated and include 'ungrammatical' telegram style sentences
  - Reman: complex sentences with three segments from literature
  - Electoral Tweets: everyday language usage from twitter users

  ➜ We train and evaluate our target classifier on all three of these very different corpora

research question
method choice

How do a naïve and a complex algorithm differ in their labelling results?

- sequence labelling is harder than nominal classification and needs context information

- compare a naïve approach without much context to a complex method

  - Hidden Markov model – takes the context of prior labels but not of tokens into account

  - Transformer – uses ???

method
Hidden Markov/viterbi

- a Hidden Markov model is trained on observations in the training data

- easily trained only with frequencies:

  - emission probabilities – compute $\frac{\text{frequency of token,tag−pair}}{\text{overall token frequency}}$ for all tokens

  - transition probabilities – compute $\frac{\text{frequency of tag}_1\text{,tag}_2\text{ bigram}}{\text{frequency of tag}_2}$ for every tag O, B and I

  - prior probabilities – compute the relative frequency of each tag as the first tag

- the best labels for a token sequence are the ones with the highest product of probabilities

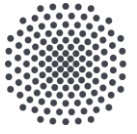- Viterbi is used to determine the labels with the maximum sequence probability

method
RoBERTa

- Transformer

- X epochs

evaluation
interpretation

- intersections between predicted and real target sequence are counted as true positives

- empty intersections are counted as false classifications

- multiple sequences mapped onto one are only counted once