

Thesis Meeting
02.06.2015
Sebastian Köpsel

Overview

- Started working with Weka
- Build an tweet to arff file builder
 - Implemented some additional features
- Increased corpus
 - I + an independent person
- Results

Working with Weka

- You got my first results, using only Bag-of-Words
- Implemented more features
 - I will show you them live
- Lots of testing, using various stopwords filters, stemmers, feature selection etc.

Arff Builder

- .arff is wekas preferred data format
- Can read twitter corpora and apply features
- Easily extendible with features

Corpora

- Added 150 more
- Independent person 150 too
- They are potentially the same, so we can compare 2 annotators
- I realized I have too much clickbaits, so I added some more “serious” twitter users, so I can balance the corpora next week

Results

- Constantly at 82 \pm 3% atm
- Test with Test set gave me 88%+ but its difficult to accomplish compatible test data...
- Big Stopword filter lower accuracy drastically
- No attribute selection is bad
- No real serious deciders yet
- Sentiment analysis not strong

Next Steps

- Searching for a method to include StanfordCoreNLPs sentence structure trees
- Build “perfect” training corpora with around 100 entries for each?
- Cleaning corpora
- More features