**Methods for sentiment analysis**

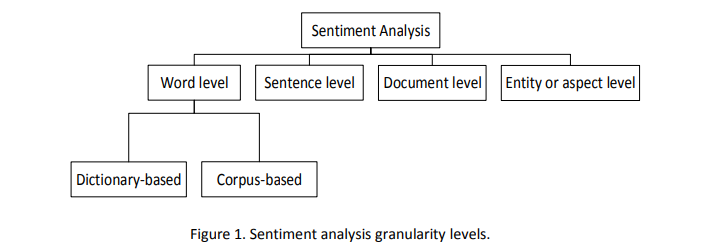
***Task****:* *Topic-Based Message Polarity Classification: Given a message and a topic, classify whether the message is of positive, negative, or neutral sentiment towards the given topic. For messages conveying both a positive and negative sentiment towards the topic, whichever is the stronger sentiment should be chosen.*

One way to analyse the sentiment of a text is to consider the **text as a combination of its individual words** and the **sentiment content of the whole text as the sum of the sentiment content of the individual words**.

* Useful Tutorial: Sentiment Analysis in R: *https://www.kaggle.com/rtatman/tutorial-sentiment-analysis-in-r*
* ML-based
* Dictionary-based (generate lists of positive and negative words and combines them into a single sentiment score. There are different dictionaries implemented.)
* Add. tailored dictionaries (customized to a specific domain, improve prediction performance compared to pure dictionaries)
* Packages:
  1. ***SentimentAnalysis****: https://cran.r-project.org/web/packages/SentimentAnalysis/vignettes/SentimentAnalysis.html*
  2. ***tidytext****: https://www.tidytextmining.com/twitter.html*

*(Dictionaries:*

* + - *AFINN-* *score between -5 and 5;*
    - *bing-* *positive and negative;*
    - *nrc-* *categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.)*



(Source: https://www.multisensorproject.eu/wp-content/uploads/2013/09/D3.3\_SentimentExtractionSocialInteraction\_20150831\_v1.0.pdf)

1. **Tumasjan**:

* Used LIWC2007, text analysis software to extract sentiment of tweets. Uses different sentiments, such as Anger, Sadness, Certainty, Achievement …
* Uses interesting visualisation of radar charts to compare different political topics with respect to each sentiment.
* Measures the average distance (via appropriate distance metric) from the mean profile per category of all politicians across all dimensions/sentiments.
* Readworthy discussion of limitations of comparable twitter data

1. **Ansari:**

* Takes into account 8 different sentiments based on different political allignments (in favour of each political party / pairwise / non / non-relevant). A score of 0-7 will result for each tweet.
* Readworthy discussion of limitations of comparable twitter data
* 2-phase-format of sentiment analysis:
  1. Collection of relevant data: based on a list of target keywords, manually, semi-automatic
  2. Extraction of sentiments: unsupervised based on dictionary / ratio of occurance of positive, negative keywords; supervised learning techniques rely on tweets that have been manually classified (i.e. hashtags) -> here only supervised
  3. Possible: Combination of lexical based and learning based approaches
* Preprocessing steps:
  1. Tokenization and filtering out of hashtags, mentions…
  2. unique tweets -> delete duplicate tweets
  3. Stopwords removal
  4. Lower case
* Feature Extraction:
  1. Term-Frequency
  2. TF-IDF
  3. Unigrams, Bigrams, Trigrams
* ML-classification algorithms for each sentiment class:
  1. SVM
  2. Decision Tree Classifier
  3. Logistic Regression
  4. Random Forest
* Long Short Term Memory: Deep learning model based on RNN architecture
* Evaluation based on:
  1. Precision
  2. Recall
  3. F1 Score

1. **Petkovic**

* Unsupervised Sentiment Classification consists of 2 parts:
  1. general or domain dependent opinion dictionary for words 🡪 expand with domain specific opinion words via semi-supervised approach
  2. classify the sentiment based on a statistical measure 🡪 Word-Level Sentiment, Aspect-Level Sentiment

1. **Devlin:**

* BERT: Sentiment Analysis with Deep Learning (advanced topic)
* <https://www.youtube.com/watch?v=-9evrZnBorM&ab_channel=YannicKilcher>
* <https://www.youtube.com/watch?v=iDulhoQ2pro&ab_channel=YannicKilcher>
* <https://towardsdatascience.com/multi-class-sentiment-analysis-using-bert-86657a2af156>
* <https://www.coursera.org/projects/sentiment-analysis-bert?action=enroll>