# Word embeddings for predicting political affiliation based on Twitter data

Ibrahim[1], Saurabh[1], Oliver[1], Venkatesh[1], Shridhar[1], Angjela[1], Shriram[1]

[1]Technische Universität Kaiserslautern

## Motivation

- Plethora of means to communicate political alignment
- Twitter, one of the main source for political opinion expression
- Quantitative analysis of German party affiliations of a Twitter user not explored

Approach proposed: deep learning based classification model using state-of-the-art word embeddings

## Data Sets and Feature Extraction

### *Dataset Collection*

- Tweets made by political figures on Twitter from the 7 biggest parties: CDU, CSU, SPD, FDP, Grüne, Linke, AFD
- Categorized data set taken from www.wahl.de/politiker
- Dividing data into train set and test set for supervised learning

### *Feature Extraction*

- Obtain vector representation of words under some similarity metrics (word embeddings)
- Word2Vec model - pre-trained on 200 million German Tweets
- Each word of the Tweet is represented as a 200-dimensional vector

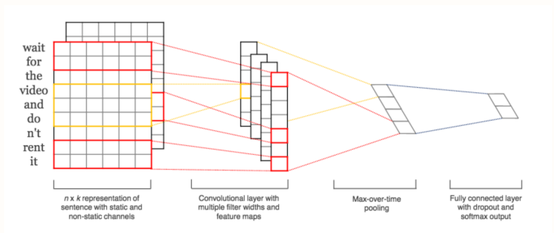


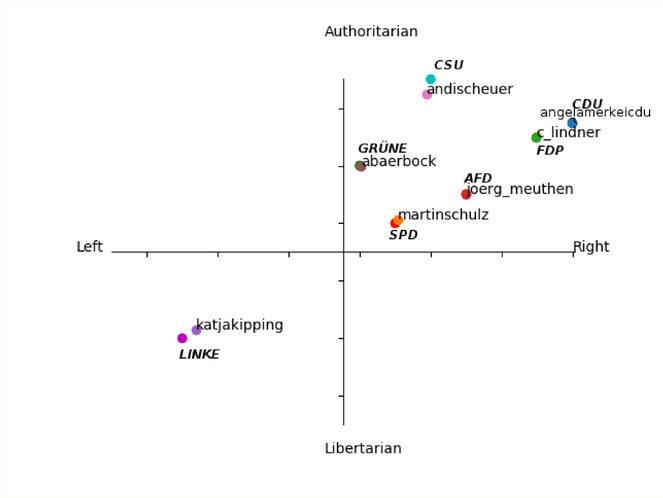

Figure 1: Model Architecture [Britz, 2015]

## Methodology

- Using a CNN to learn features from different word-lengths for classification
- Training done by feeding data using one-hot-encoded labels
- Cross Entropy to define the loss and also to optimize the network

## Results

- Accuracy
  - per single tweet for person low accuracy 0.55
    - Insufficient number of words
    - Retweets
    - Non-German messages
    - Low numbers of Tweets.
  - per multiple tweets high accuracy 0.817
- Visualization through a political compass
  - Users appear close to their respective political party
  - Parties with similar Ideologies appear close to each other




## Conclusion and Future Work

- CNN is able to nicely separate political figures concerning party affiliations
- CNN gives better results when used for short sentences such as tweets
- Further research may tackle:
  - How the analysis would be affected if the embeddings were to be taken from intrinsically political data samples
  - Consider RNNs or other topologically different architectures

## References

[1] Maneesh Bhanda, Dan Robinson, and Conal Sathi. Text classifiers for political accuracy that they report, it is stated that the ideologies. 2009.

[2] Felix Biessmann, Pola Lehmann, Daniel Kirsch, and Sebastian Schelter. Predicting political party affiliation from text. 2017.

[3] Raviv Cohen and Derek Ruths. Classify-ing political orientation on twitter: It's not easy!

[4] Yoon Kim. Convolutional neural networks for sentence classification

[5] Britz, D. Implementing a CNN for text classification in Tensorflow. 2015.