Word embeddings for predicting political affiliation based on Twitter data

Ibrahim Abdelaziz[1], Oliver Berg[1], Angjela Davitkova[1], Venkatesh Iyer[1], Shriram Selvakumar[1], Kumar Shridhar[1], and Saurabh Varshneya[1]

[1]Technische Universität Kaiserslautern

January 29, 2018

1 Abstract

Twitter as one of today's biggest social media platforms allows political figures to express their thinking easily and with a concise message. We propose a generic way of classifying political affiliation based on Twitter posts. This involves Word2Vec vector representations of the input data and utilizes pre-trained embeddings for the German language. With this we have shown to be capable of insightfully position German political figures in the political spectrum.

2 Introduction

Social media platforms like *Twitter* allows people of interest to communicate their personal opinion, and as such e.g. indicating political alignment, through a comprised message being only a few hundred character long. This yields broad potential to characterize personality traits such as political affiliation on.

Generally speaking, political motives were shown to be consistently predictable with an accuracy better than chance already [Biessmann et al., 2017]. This paper therefore proposes a *deep learning* based classification model together with *word embeddings* [Pelevina1a et al., 2016]. This allows a later analysis to find interesting constellations within the (German) political spectrum. We leverage word embeddings to represent words in context. We thereby restrict ourselves to pre-trained models. Subsequently, a convolutional neural network (CNN) holistically classifies the Twitter profile by assigning each Twitter message a separate party label and combining these into a complete class score.

3 Related Work

Today, sentiment classification is mostly done using recurrent- or convolutional neural networks as described in [Kim, 2014]. The presented approach uses a basic CNN trained on pre-trained word vector representations and does only little hyperparameter tuning to already achieve compelling results in question classification.

In connection to the given focus of working on Twitter data, [Cohen and Ruths, 2013] further introduces interesting questions concerning applicability of classification onto real data outside the training samples. It depicts the validation process as being prone to optimistic interpretations of the result when overlooking problems in latent attribute inference. This also suggests a critical view on this paper's final analysis results.

In addition, [Sug, 2007] motivates context- and time dependencies within political data, which again makes analyzing an overall corpus of Twitter data a broadly connected issue.

4 Methodology

Formally, this paper proposes work on classifying Twitter data of political figures and picturing tendencies and dependencies within the data. As such, the following methodology is being applied.

4.1 Dataset

The dataset used in this approach was constructed of German politicians' Tweets posted on their Twitter accounts. 1000 German politicians' profiles with attributes "name of person", "political party" and "twitter username" were retrieved from the website "https://www.wahl.de/". To reduce noise, only politicians belonging to the seven major political parties with respect to parliament activities, namely "CDU", "CSU", SPD", "FDP", "GRÜNE", "LINKE" and "AFD", are considered.

After the filtering stage this leaves a list of around 700 politicians, for which up to 1000 of the most recent Tweets are gathered using the Twitter API. To equally compare parties of different sizes, each party is restricted to a total of 12000 Tweets, where each politician contributes approximately the same number of tweets.

To prepare the Tweets for the training step, each Tweet was preprocessed by first removing URLs, special characters, user names, and mentions. Then all characters are masked and put to lowercase.

To represent the preprocessed Tweets text into numerical values that can be used to train our model a pre-trained Word2Vec model is used [Mikolov et al., 2013] - pre-trained on 200 million German Tweets [Cieliebak et al., 2017] - which represent each word of the Tweet as a 200-dimensional vector.

4.2 Classification

The 200-dimensional word vectors are fed to a CNN model for classification. The employed CNN architecture, shown in Figure 1, is based on [Kim, 2014] which uses its model for sentiment analysis specifically.

The first layers embeds words into low-dimensional vectors. Subsequent layers performs convolutions over the embedded word vectors using multiple filter sizes; sliding over 3, 4 or 5 words at a time. Max-pooling transforms the result of the constitutional layer into a long feature vector, dropout regularization is added and the result is classified using a final softmax-layer.

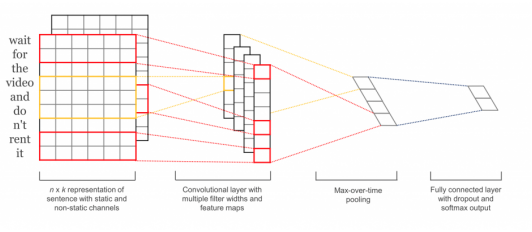

Figure 1: Model Architecture

The overall classification process can be sub-divided into two elementary parts: First we **feed input data** to the neural network to then **classifying the (test-)data** for their respective classes.

In order to feed input data: For each party a raw text collection of all Tweets is considered. For each line - containing a single Tweet - a <PAD> tokens is appended to achieve a fixed size length of size 35 which allows efficient batch processing. A vocabulary built on the complete corpus of existing words within the data maps each word to an integer between 0 and 109933 (number of existing words) for faster indexing. Note that each Tweet is now represented as a vector of integers only. Using the pre=trained Word2Vec model, each Tweet can then be represented as a Matrix $M \in \mathbb{R}^{35\times 200}$.

For classifying the user Tweets to the correct political party, we then feed the batches of twitter data along with their correct political party one-hot-encoded labels and train the above defined CNN. To optimize the network we use cross entropy loss defined as

$$H_{y'}(y) = -\sum_i y'_i \log(y_i)$$

where y is our predicted probability distribution, and y' is the true distribution (the one-hot vector with the true-class party labels).

4.3 Analysis

[...]

5 Quantitative Analysis

For a quantitative analysis of the results, every Tweets are taken as sentences and a sentence vector of the same dimension as the word vectors used for training (i.e. 200 dimensional vector) is constructed. In doing so, the Tweet is tokenized into words and every word's corresponding token vectors are taken from the word embedding space. A sentence embedding vector is constructed by adding the word vectors and dividing them by the number of words. Note that only those Tweets for which all its words were present in the word embedding space are considered and that any Tweets with *UNK* token were ignored for a cleaner analysis. This effectively removes any Tweets that were not completely related to political topics such as simple link- or hashtag-forwarding.

The sentence embedding was used to visualize the vectors based on their 200-dimensional vector representations. The initial results were generally pretty mixed up with no clear distinct clusters for any specific party.

We found several reasons for this: Since a user Re-Tweets other parties Tweets with his opinions, it could be that subjects get mixed rather quickly across parties. Another reason for such mixed clusters could be the fact that any collection of party-members is typically tweeting about similar issues with some in favor of and some against some specific proposition, which may lead to strictly disconnected sentiment groups within the same party.

As such, a direct embedding as described earlier did not prove to be a good measure to visualize the result. This motivates our usage of the "political compass" [Pol, 2017] for visualizing the result further.

For the political compass representation, we use a four way graph plot with attributes of "Left-wing" and "Right-wing" on the X-axis as well as "Authoritarian" and "Libertarian" on the Y-axis. On the four axis, we used a one hot encoding to plot all the seven classes on the compass. This way we can see all the political classes plotted at once.

5.1 Visualization

Now, we need to plot all the users on the compass to see which users corresponds to what cluster by:

1. Passing all of a user's Tweets to the classification model as specified above in the model architecture section
2. Obtaining the class classification and probability score per Tweet concerning all seven classes
3. Averaging all probability distribution to get one vector representing the distribution over all classes
4. Plotting the user on the political compass to visualize the results and to compare whether the user has been classified to it's correct class.

5.2 Inference

The subsequent visualization was created as described above.

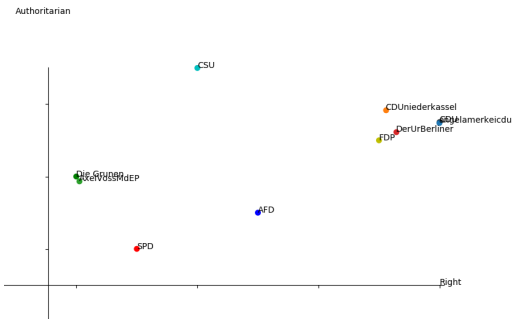

Figure 2: Plotting on Political Compass

From this depiction within the political compass, we can derive the following:

1. Users generally appear close to their respective political party; e.g. Angela Merkel is very near to CDU

2. The FDP-class is plotted close to the CDU-class. Both party's ideology are traditionally quite similar, which matches the visualization

3. The user "AxelVossMdEP" from the party CDU was found to appear very close to Die Grünen. In our context, this indicates a strong affiliation towards Die Grünen instead of CDU. Upon analyzing the data, we observed a common occurrence of Tweets in the English language, which will have led to a huge number of <UNK> token generation, hence the classification may not be correct. After manually cleaning the data, we were able to see the user's plot moving to the correct class CDU.

6 Results

Classification to a specific party based on a single tweet is a difficult task, even for humans. We, through our experiments comprehend the same as our results on individual tweets gave a low classification accuracy of 55 percent. However, taking majority voting over a set of tweets for a specific user led to a considerable increase in the classification accuracy depicted below:

party	**users**	**tweets/user**	**accuracy**
AFD	15	500	86.67
CSU	16	500	87.50
FDP	22	500	81.81
Grüne	42	700	66.67
Linke	28	700	89.28
SPD	85	500	69.40
CDU	67	500	76.20

7 Conclusion

As we have shown, a comparably simple convolutional neural network is able to nicely separate political figures concerning party affiliation. Interesting findings like single entities more closely connecting to general party-affiliated language or specifically different language have been pointed out.

As the underlying word embeddings are taken from the German-language Wikipedia dump, we are currently restricted to German-language Tweets as well as to overall German-language features. This does suffice for general classification purposes, but poses the additional question of how the analysis would be affected if the embeddings were to be taken from *intrinsically political* data samples. Also, our approach primarily focuses on CNNs and proves them to be efficient already. For future work, it would be intriguing to compare the capabilities of RNNs or other topologically different architectures.

References

[Sug, 2007] (2007). Ideology classifiers for political speech.

[Pol, 2017] (2017). German general election 2017.

[Biessmann et al., 2017] Biessmann, F., Lehmann, P., Kirsch, D., and Schelter, S. (2017). Predicting political party affiliation from text.

[Cieliebak et al., 2017] Cieliebak, M., Deriu, J., Egger, D., and Uzdilli, F. (2017). A twitter corpus and benchmark resources for german sentiment analysis. *SocialNLP 2017*, page 45.

[Cohen and Ruths, 2013] Cohen, R. and Ruths, D. (2013). Classifying political orientation on twitter: It's not easy!

[Kim, 2014] Kim, Y. (2014). Convolutional neural networks for sentence classification.

[Mikolov et al., 2013] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

[Pelevina1a et al., 2016] Pelevina1a, M., Arefyev, N., Biemann, C., and Panchenko, A. (2016). Making sense of word embeddings.