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

Introduction

- Plethora of means to communicate political alignment
- Twitter, one of the main source containing message specific to general public
- Aim is the quantitative analysis of party affiliations of a user
- We propose a deep learning based classification model using state-of-the-art word embeddings

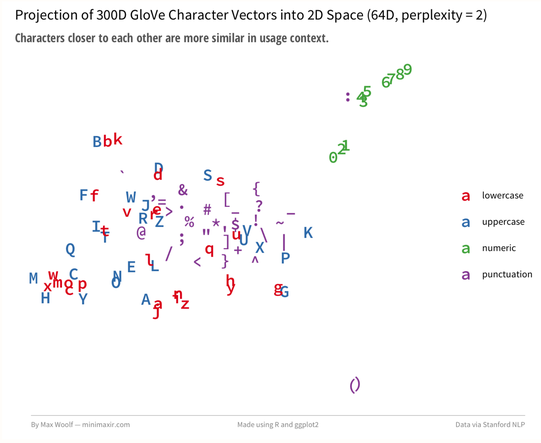
Data Sets and Feature Extraction

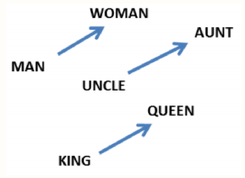
Dataset Collection

- Tweets made by political figures on Twitter
- Categorized data set taken from www.wahl.de/politiker
- Dividing data into train set and test set for supervised learning

Feature Extraction

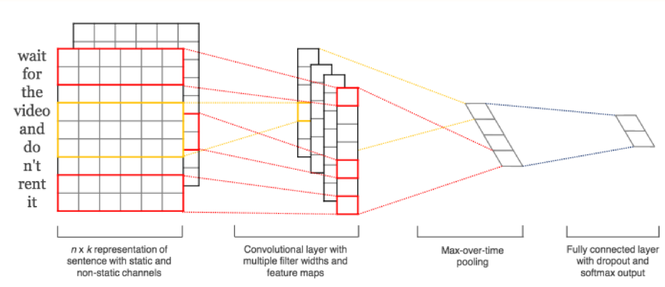
- Obtain vector representation of words under some similarity metrics (word embeddings)
- WordtoVec using Continuous Bag of Words and Skip Grams
- GloVe embeddings performed on global word-word co-occurrence statistics from a corpus





Recent Developments

- Existing research [2] tackle the text classification problem by techniques like SVM, SVD and LSM
- Approaches solely considers political affiliations in America, which is relatively easy problem
- Many research papers compare different classifiers [1], and do not propose a well-developed solution
- Sentiment classification is mostly covered using recurrent- or convolutional neural networks [4]
- Classifiers cannot be used for classifying users outside the training data.[3]



Proposed Methodology

- Using a CNN to learn features from different word-lengths for classification
- Use bi-directional LSTM, which caters the need to learn the context in temporal space
- Idea is to use a combination of above approaches in an end-to-end learning

Analysis of Results

- Quantitative analysis of result under various metrics like F1-score, cross-entropy logarithmic loss
- Visual analysis of embedding space and results using T-SNE or PCA

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