

Sentiment Analysis of SOTU speeches

Ananya Mukherjee

March 3, 2018

1 Introduction

In this report I look into the State of the Union (SOTU) speeches from 1901 to 2017. Particularly we try to assign each speech a ‘sentiment score’, and then try to draw some conclusions from the results.

2 Method

The AFINN dataset[1] a list of English words rated by hand for valence or sentiment with an integer between minus five (negative) and plus five (positive). It has 2477 words. For each speech, we replace each word by its AFINN score. For words not in the dataset, we use a score of 0. Note that stopwords removal is not necessary, since the AFINN dataset does not have scores for stopwords, hence they are automatically removed from consideration.

For each speech we then compute the average sentiment score and the average score, when not considering words not present in the dataset. The results are described in the section below.

3 Results

Fig 1 shows the average sentiment score including zero-score words, while fig 2 only included non-zero score words in the average. Both graphs have similar trends and have been colour coded blue or red to indicate Democrat and Republican presidents respectively. Different shades of red and blue are used to demarcate between consecutive Democrats and Republicans. At the

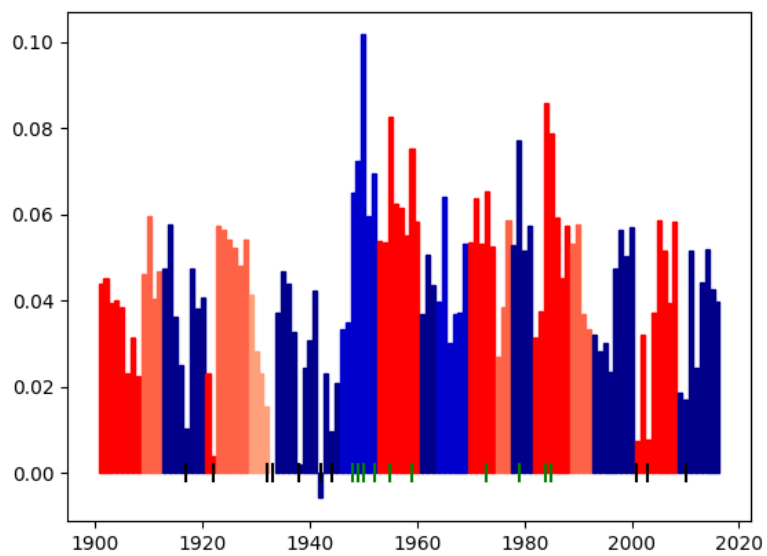


Figure 1: A boat.

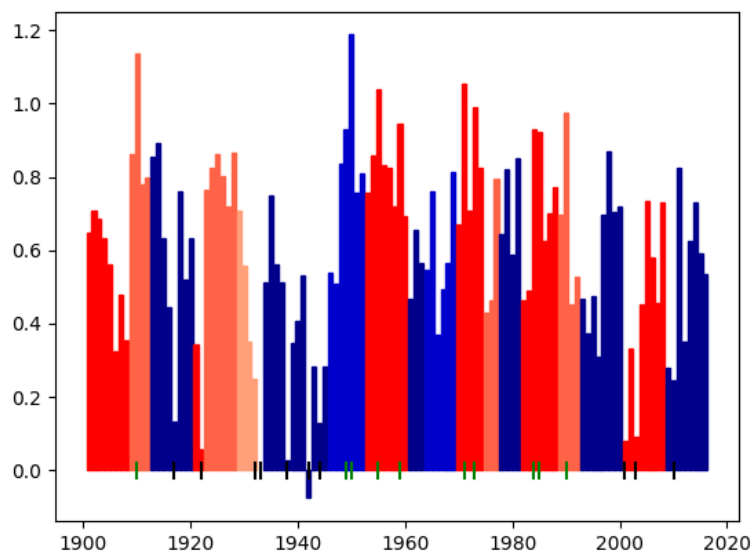


Figure 2: A boat.

base of each bar, for some years there are black and green markers. Green markers denote the top 10 years and the black markers denote the bottom 10 years, sentiment-score wise.

Note that the first half of the 20th century and the 21st century years have somewhat low sentiment scores. The 2nd half of the 20th century has high scores. This is also borne out by the green and black markers, which show that the highest scores are between 1950 to 2000 and the lowest scores are in the other years. This phenomenon is explainable by the occurrence of the Great Depression, the two World Wars and the more recent 9/11, Iraq War and the Great Recession. Note that in the years corresponding to these events, we have a black marker, indicating lowest sentiment scores.

Thus, sentiment scores of SOTU speeches seem to correlate well with events in those years.

References

- [1] F. Å. Nielsen. AFINN, mar 2011.