A NLP Approach to Understanding Variation in Political News

Stephen M. Lee

July 23, 2019

1 Introduction

The 2016 United States Presidential Election, to many, raised questions as to the reliability of their news sources [SOURCE NEEDED]. Since then, many social media companies and other institutions have begun public campaigns to combat the perceived threat from fake news, with arguably limited results [SOURCE NEEDED]. With a goal of news article source identification, I scrape several thousand news articles from Fox, Vox, and PBS. By some estimations [SOURCE NEEDED], these three represent distinct categories of news: Fox is often considered extreme right opinion (i.e. conservative); Vox is considered extreme left opinion (i.e. liberal); and PBS is considered center primary source news. I first calculate the top word, 2-gram, and 3-gram frequencies to better understand the dataset, and then train a Bidirectional LSTM neural network with pretrained embeddings to classify the news source.

2 Data

3 Predictive Models

4 Analysis

5 Conclusion and Discussion