

KU Leuven

Modern Data Analysis

Report

Yitong Li

r0867865

MSc Statistics and Data Science

Faculty of Science

August 2022

1 Overview

This research will use various NLP techniques to analyze the speech of Barack Obama during his two consecutive presidential terms from 2009 to 2017. This project aims to gain insights into the major topics covered in the selected speeches and investigate how external factors may influence the frequencies of different topics in Obama's speech.

2 Data Preprocessing

In data preprocessing part, I applied the spaCy package to clean the speech content. SpaCy can parse and tag the given document after tokenization. After using the spaCy, we can obtain the tokenized texts and removed the words labeled by 'PERSON', 'ORG', and 'DATE'. For the sentiment analysis, we should keep the parts tagged as 'ALPHA' and not tagged as 'STOP'. The reason is that the words like adjectives have an impact on sentiment analysis. For topic modeling, it is necessary to remove words such as 'ADV', 'PRON', and 'CCONJ'. In addition, we can count the frequency of individual words. Some highest frequency words, like 'american', 'america', and 'everyday', were not removed after the preprocessing using spaCy, but they make no sense in topic modeling. Hence, we need to delete these words with the lowest frequency to improve the accuracy of topic modeling.

3 Sentiment Analysis

In this section, I used the TextBlob to do a sentiment analysis. Basically, sentiment analysis is the process of determining the attitude or the emotion of the text, i.e., whether it is positive or negative or neutral. The TextBlob.sentiment function returns the polarity property which lies in range from -1 to 1. From figure 1, we can notice that all most all speeches indicate a positive polarity. However, the polarity scores are smaller than 0.5 and the most polarity scores lie in [0.05, 0.2]. Hence, we can conclude that the emotion of the speeches is approximately neutral

4 Topic Modeling

In this section, the LDA technique is used to identify the main topics of Obama's speeches and the topics distribution. To determine the optimal number of topics, we can use the perplexity, where a low perplexity indicates the LDA model is good. Figure 2 shows that when the number of topics equals 7, the perplexity is the lowest. Hence, I selected 7 as the number of topics for the topic modeling and print the top words. The 7 topics could be merged into several speech themes, assigned with the meaningful names: economy, health care, war and violence, and foreign relation. Finally, I loaded the topic distribution for each speech texts and generated the time series plot to visualize the topic distribution, to discover the effect of extra variables on

the topics.

5 Interpretation

As shown in figure 4 (Gallup, 2017), President Obama's approval rating rose relatively quickly from 41% in August 2011 to 53% in December 2012. During this period, we can notice that the primary topics of his speeches were war and jobs, with the topic distribution of war reaching over 50% and the topic distribution of jobs reaching almost 68%. Based on the event timeline (Barack Obama Event Timeline | The American Presidency Project, n.d.), we can see that after Al-Qaeda leader Osama bin Laden, who was responsible for the 9/11 attacks, was killed by US forces in Pakistan, a record number of Americans opposed the war. Under the pressure from lawmakers and the public, Obama decided to sizably reduce U.S. forces in Afghanistan, announcing on June 22, 2011, that 33,000 troops would be withdrawn from Afghanistan by the summer of 2012 (Laub, 2017). In October of that year, President Barack Obama announced another strategy to end combat missions in Iraq and remove all U.S. troops by the end of 2011. His decisions and speeches on the war may have been one of the factors that boosted his approval ratings.

In addition, Obama's job-related activities, such as the introduction of the American Jobs Act in September 2011, may also help to enhance his approval ratings. According to a report by Pew Research Center (2020), Obama's job approval ratings improved considerably in December 2012, corresponding to greater optimism of the public toward the economy. While not all respondents considered the economy to be in good shape, the number of people describing economic conditions as poor dropped to 35%, the lowest since January 2008. 22% of respondents believed that plenty of jobs were available, which is 10% greater compared to early 2010 and the highest since 2008. The corresponding trend can also be seen in the unemployment graph (figure 5).

6 General Conclusion

Based on the above analysis, I believe that approval ratings may be one of the factors influencing Obama's speech. When his approval rating was at its lowest point in August 2011, his speeches focused on areas of significant public concern (war and job), which was likely to raise his approval rating and establish the groundwork for a smooth re-election in his second term.

7 Reference

Gallup. (February 21, 2017). Do you approve or disapprove of the way Barack Obama is handling his job as president? [Graph]. In Statista. Retrieved June 07, 2022, <https://www.statista.com/statistics/205284/obama-job-approval-rate-by-the-american-public/>.

Barack Obama Event Timeline | The American Presidency Project. (n.d.). The American Presidency Project. <https://www.presidency.ucsb.edu/documents/barack-obama-event-timeline>.

Laub, Z. (2017, May 1). The U.S. War in Afghanistan. Council on Foreign Relations. <https://www.cfr.org/timeline/us-war-afghanistan>.

Bureau of Labor Statistics. (February 10, 2022). Number of unemployed persons in the United States from 1990 to 2021 (in millions) [Graph]. In Statista. Retrieved June 07, 2022, <https://www.statista.com/statistics/193254/unemployment-level-in-the-us-since-1990/>.

As Fiscal Cliff Nears, Democrats Have Public Opinion on Their Side. (2020, July 28). Pew Research Center - U.S. Politics & Policy. <https://www.pewresearch.org/politics/2012/12/13/as-fiscal-cliff-nears-democrats-have-public-opinion-on-their-side/>.

8 Figures

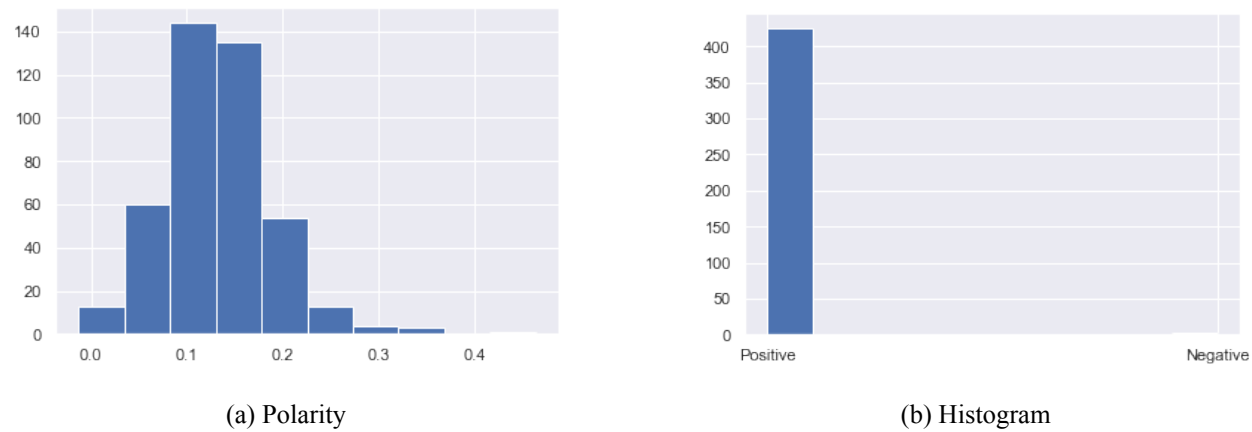


Figure 1: Sentiment Analysis

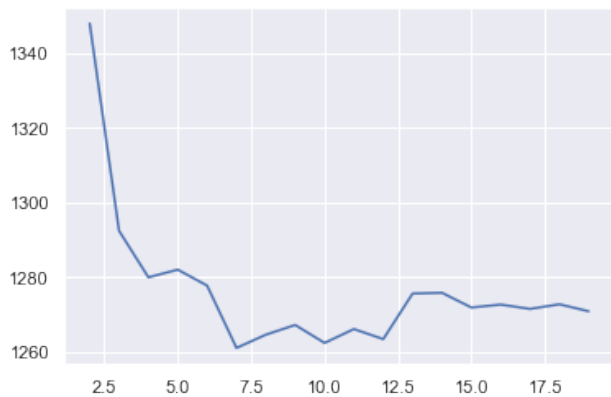
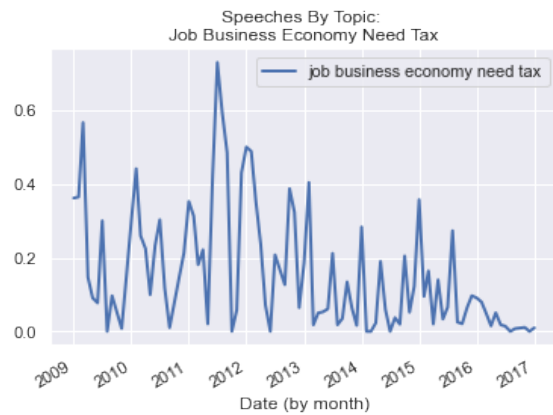
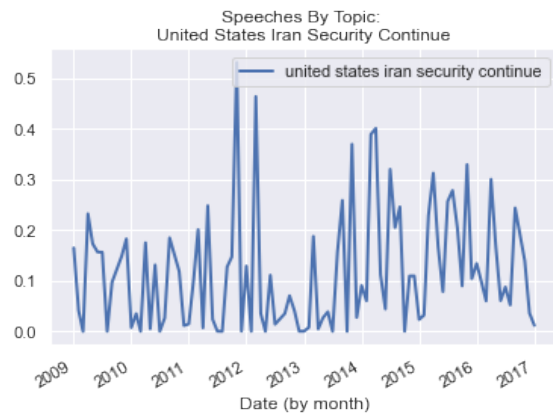


Figure 2: LDA



(a)



(b)

Figure 3: Sentiment Analysis



Figure 4: Obama Presidential Job Approval 2009-2017

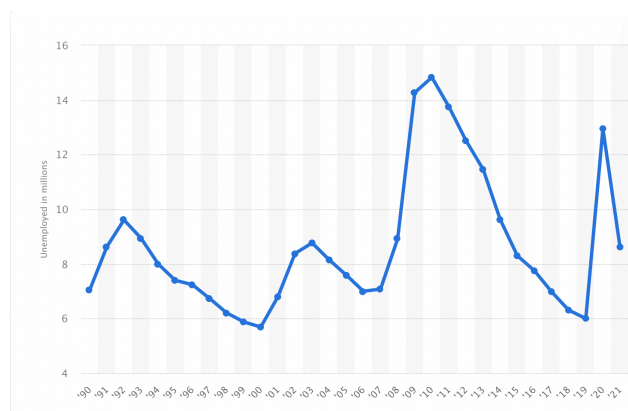


Figure 5: Unemployment Rate from 1990 to 2021