Assessing U.S. Politics Through 234 Years of Presidential Speeches

Project Motivations

Over the course of the history of the United States, there have been substantial changes not only in regards to the nation's sentiment towards varying issues, but also within the national lexicon chosen to express it. While American History does an effective job at revealing dark periods in regards to attitudes towards varying countries, ethnic groups, and others, we sought to quantify said sentiments and their respective changes by utilizing current developments in Natural Language Processing. To do this, word embeddings and sentiment analysis techniques were implemented on a dataset containing every presidential speech delivered over the course of the United States' 234 year history. This research was inspired by that of Garg et al (2018), who examined the evolution of gender and ethnic stereotypes in America over the course of a one-hundred year period.

Review of Literature

Word Embeddings/Sentiment Analysis Literature

In recent years, machine learning and natural language processing methods have grown increasingly popular in both academia and industry, due in part to an increase in availability to corpora to help train their models. One popular method for text analysis is word embeddings. In word embedding models, each word in the corpus exists in a high-dimensional vector space, and the geometric relationships between the words are telling of semantic relationships, allowing models to capture deeper subtleties and nuances of the English language, such as analogies (Garg et al., 2018). For example, as seen in Figure 1 below there exists a two-dimensional space with the words "king", "queen", "man", and "woman". By analyzing the geometric relationships between these vectors, one can perform math with them (i.e. King - Man = Queen).

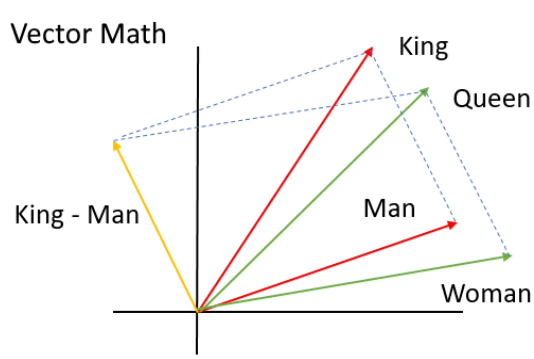

Figure 1: Vector representation of King/Queen/Man/Woman (Talreja, 2020)

Another increasingly used method for textual analysis and natural language processing is sentiment analysis, as it can be applied to a number of domains such as the financial markets, politics, news, and social movements (Wang et al., 2012). Sentiment Analysis is the process by which emotional polarity is assigned to a piece of text - with a piece of text being assigned as either positive, negative, or neutral (Medhat et al., 2014).

Sentiment / Embeddings Analysis to Track Cultural/Political Trends

A number of studies have explored the utilization of natural language processing techniques to track trends in the political and cultural landscape in the United States. Computational Analysis on congressional speeches and presidential communications has found that when examining sentiment around immigration, it has become much more positive on average, despite the data revealing increasing polarization between political parties. Additionally, researchers found trends and deviations in the vernacular that was used to describe the varying ethnic groups that were immigrating into the country (Card et al., 2022). Other research has analyzed US Census data and 100 years of text data to explore and quantify the changing nature of gender and ethnic stereotypes, finding that these machine learning methods can in fact be reliably used to complement historical analysis (Garg et al., 2018).

Sentiment analysis is also prominently used when examining political discourse, cultural impact, and attempting to track the collective attitude either in real-time or over the course of history. For example, in a study that tracked tweet data in real time over the course of the 2012 election, it was found that tweet data was very responsive to events happening in the physical space, with sentiment online changing rapidly in response to campaign and debate events (Wang et al., 2012).

Questions

1. How have word similarities changed over time when considering ethnic/gender based terms?
2. Is there any relationship between the terminology used and the general sentiment during that time period in the United States?
3. How does sentiment/word embeddings vary when considering different subsets of the data (i.e. Political Party, Era in American History, etc.)?

Method

Data Source

For the purposes of this research we chose to work with the "United States Presidential Speeches" dataset from Kaggle. This dataset contained transcripts of every presidential speech from April 30, 1789 (George Washington's Inaugural Address) to September 25, 2019 (Press Conference with Donald Trump). In addition to the transcripts, the dataset also contained information pertaining to the President who gave the speech, the political party of which they were affiliated, the speech title, as well as a short summary of each speech.

Preprocessing

In order to work with this data in a meaningful way, a great deal of preprocessing needed to be done. Firstly, loading the data returned all columns as objects, so preprocessing needed to be conducted on date to work with it as a datetime object. The year was then extracted and added as a feature to simplify later time-series analyses. Additionally, all speech transcripts were then casted as string objects so that more preprocessing could be performed.

Once the dataset was prepped, work shifted towards prepping the transcripts for future work with our Word Embeddings/Sentiment Analysis models. The first step of this process was to remove punctuation from the transcripts, as it will help in regarding each word in a text equally. The second step in this process was removing stop words from each speech, as this would remove low-level information from each text and allow deeper focus on more important information. Each respective transcript was then tokenized, or broken down into a sequence of its singular word elements, and finally the data was ready to train our models.

Word Embeddings Modeling

The goal of the word embedding models can be split into three major steps. The first step was to subset the speeches using different splits. We wanted to compare results of word embeddings and analyze how the results compare to our preconceived understanding of the topic. The subsets that we used were 20-year splits on all speeches regardless of party and 40-year splits based on political party. The second step was to create a set of models using the subsets of data. We then used those trained models and a set of pre-selected keywords to find the most similar words to each keyword for each subset of data.

The method that was used for the modeling was Doc2Vec from the Python package Gensim. Doc2Vec is a method similar to Word2Vec, but expands on it by representing an entire document as a vector rather than a word. Doc2Vec has methods to easily convert a trained model to Word2Vec, so features from both types of models can be used. We trained our Doc2Vec models on the subsets of data and analyzed the similarity between the speeches. We also convert the models to Word2Vec to find the most similar words to each keyword.

Sentiment Analysis Modeling

There were two main focuses for evaluating the sentiment of the presidential speeches, comparison of sentiment over time and comparison of sentiment between parties. To quantify the sentiment for each speech, we used the Spacy python package to create an NLP model that would be used to analyze each piece of text. We then used the Sentiment Intensity Analyzer from the NLTK package to generate the sentiment scores for each speech. The compound sentiment score was then added to the dataset as a new column.

The next step was splitting the data by party and by era. Using the party-split data, we calculated the running average sentiment score, to be visualized later, within each party. Once the running averages had been calculated, all of the parties' tables were re-aggregated into one dataframe. We then split the data into five different eras in order to assess differences in sentiment for each party across eras. To aid in this analysis, we created various visualizations that plot line graphs of each party's running average sentiment over time. These visualizations were created for each individual dataset as well as each era. In addition to these visualizations, we calculated the average sentiment of each party within each era in order to compare parties within an era as well as a particular party across different eras.

Results

Word Embeddings

Using the presidential speech data split into 20 year increments, there were 11 word embedding models trained (one model for each 20 year split from 1789 to 2019). By splitting the data in this way, we were able to identify how similar words to specific keywords have changed over time based on these trained word embeddings and cosine similarities. A trend we identified is that similar words change based on era and what events were happening in American history during that time. The keywords we chose reflected a wide range of topics from war to the economy to political ideology and more.

When looking at the keyword “enemy”, we found that different words had high cosine similarity scores to it at various points in history. In the early 1800s the terms “Canada” and “Canadians” were similar to “enemy” due to conflicts regarding the War of 1812 where several battles were fought on Canadian land. During the World War II time period, terms like “Mindanao,” “naval,” and “fleets” were associated with “enemy” because of the U.S. military’s involvement in the Battle of Mindanao as well as the heavy U.S. naval build up over this time. In the 1980s and 1990s, the name “Saddam Hussein” and the word “dictator” were related to “enemy” as a result of Hussein’s controversial reign as Iraqi president that had heavy implications on the United States. Finally, “Laden” in reference to Osama bin Laden was also a similar term to “enemy” in the 2000s in the aftermath of the 9/11 attacks.

Enemy Similar Words Over Time		
Word	**Time Period**	**Cosine Similarity**
“Canada” / “Canadian”	War of 1812	.741 / .722
“Mindanao”	Battle Fought in Mindanao, Philippines in WW2	.79
“Naval”	WW2	.713
“Saddam” / “Hussein” / “Dictator”	During Hussein’s Iraqi Reign	.697 / .603 / .614
“Laden”	9/11	.61

Figure 2: “Enemy” Similar Words Over Time

Another keyword that produced intriguing outcomes was "slavery." The manner in which the term has been addressed has undergone significant changes depending on whether it was legal or not. Before the ratification of the 13th amendment, words such as "instrumental," "accepts," and "participating" had high cosine similarity scores with "slavery." However, as the discussion around "slavery" became more negative, phrases such as "miseries," "extreme," and "cruelty" became more closely related to "slavery" in the last two decades.

Slavery Similar Words Over Time		
Word	**Time Period**	**Cosine Similarity**
“participating”	Before 13th Amendment	.83
“accepts”	Before 13th Amendment	.754
“instrumental”	Before 13th Amendment	.747
“extreme”	Past 20 Years	.753
“miseries”	Past 20 Years	.73
“cruelty”	Past 20 Years	.725

Figure 3: “Slavery” Similar Words Over Time

America has a rich history, so it is interesting to see how presidential discourse about a variety of topics has changed throughout the course of over 200 years. While “enemy” and “slavery” are two good examples of this changing language, other keywords also produced relevant results in our models. The word “war” returns typical “war” related terms like “conflict,” “siege,” and “commanded,” but also returns a more unique term with “sputnik” in reference to the Cold War and the Space Race. American laws and policies have also undergone significant changes over time. As a result, searching the word "illegal" reveals related terms that reflect highly debated laws during different eras. For instance, during Prohibition in the 1920s and early 1930s, terms such as "liquor" and "rehabilitation" were commonly associated with "illegal," but more recently, terms such as "undocumented" and "aliens" have become more similar in regards to discussions about immigration. All in all, our results show that word embeddings reflect important American historical events and political climate.

Sentiment Analysis

The first analysis conducted using the sentiment scores, was assessing whether or not there was a correlation between party affiliation and the sentiment of a given speech. After creating a new “Party_id” column as a numerical representation of each party, the correlation coefficient found between this column and the “Sentiment Score” was -0.0664, meaning that there is almost no correlation between the party affiliation of a president and the sentiment of their speech. From this, we can conclude that no particular party consistently gives more positive, or negative, speeches than the other parties.

Next, we calculated the average sentiment for each party within each era (Figure 4). At the beginning of the country’s independent history, all of the presidential speeches were overwhelmingly positive, with every party having an average sentiment of 0.95 or higher during the first era. A notable change moving to the second era is the drastic decrease in the sentiment for the Democratic-Republican party. There were actually no presidents elected from this party during this era, however, there was a statement from John Quincy Adams before the Supreme Court ahead of a case that caused this decrease in the average sentiment. The Democratic and Republican parties emerged as dominant political voices during the second era, although the sentiment for their speeches during this period was more negative than that of the other parties. From the third era until today, there have only been presidents from these two parties and their average sentiment scores have been very positive as well, with the lowest average coming in the fourth era for the Democratic party. This is largely due to their dominance of the presidency during the Great Depression and the post-depression era through the early 1950s. In each of the third, fourth, and fifth eras, whichever party held the presidency more often held a lower average sentiment score than its counterpart.

Party	*Average Sentiment*
Era 1 (1789-1830)	
Unaffiliated	0.9744
Federalist	0.9934
Democratic-Republican	0.9587
Democratic	0.9981
Era 2 (1831-1872)	
Unaffiliated	0.9907
Whig	0.9996
Democratic-Republican	0.6783
Democratic	0.8308
Republican	0.7614
Era 3 (1873-1914)	
Democratic	0.9035
Republican	0.8898
Era 4 (1915-1956)	
Democratic	0.7886
Republican	0.9223
Era 5 (1957-2019)	
Democratic	0.9187
Republican	0.8543

Figure 4: Average Sentiment by Political Party Over Time

The last step of our sentiment analysis was to visualize the sentiment of each party over time. To do this we decided to create a line plot for each era where each line represents the different parties. To do this, we calculated the running average sentiment, within each party, for all of the speeches. From these visualizations, we were able to observe specific periods where the running average sentiment dropped significantly relative to the other speeches given during that era. While all of the five eras did not produce particularly interesting results, there were a few instances where a party's running average seemed to dip. As seen in Figure 5, the Republican party sees a very large drop in sentiment during Abraham Lincoln's presidency. This could be explained by the major events occurring during this period, namely the Civil War. In a similar fashion, we see the Democratic party's sentiment take a dip during the fourth era (Figure 6). The drop is less obvious than that of the Republican party in the second era as it is somewhat stabilized by years of speeches from previous Democratic presidents. The dip we see during the fourth era for the Democratic party occurs during the Great Depression and the years that followed.

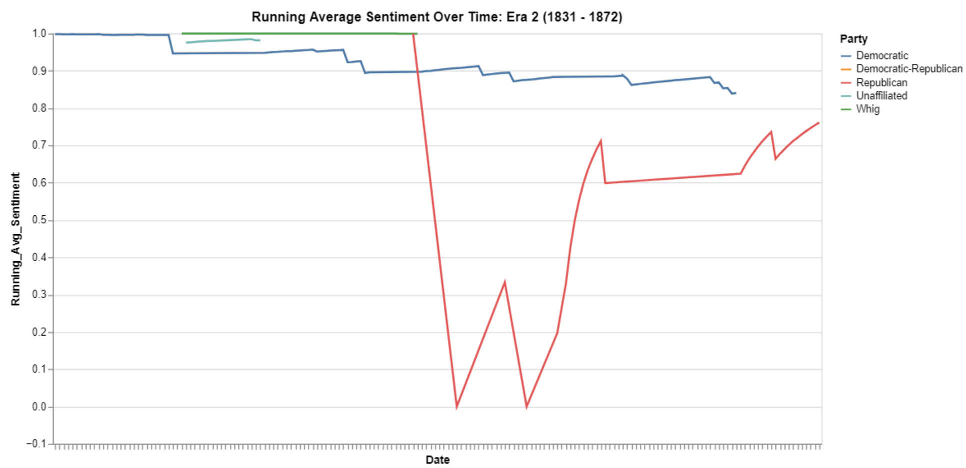

Figure 5: Running Average Sentiment Over Time: Era 2

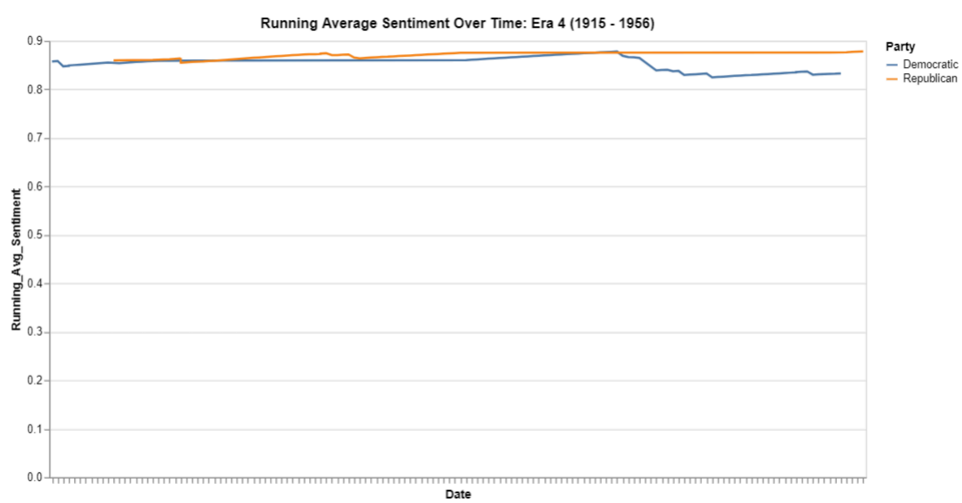

Figure 6: Running Average Sentiment Over Time: Era 2

Using the various calculations and visualizations, we can begin to form an idea of how politics in the United States have evolved over time. We found no correlation between a president's party affiliation and the sentiment of that speech. This sentiment is determined by the content of the speech, which largely focuses on the events and issues that are prevalent at any given time. In general, presidential speeches have been mainly positive, with an average sentiment score of

0.8834 for the entire dataset. Despite this, there are periods where the average sentiment will dip, often coinciding with major events in history such as the Civil War or Great Depression.

Discussion / Conclusions

After looking at the results from both our word embedding models and sentiment analysis, we can conclude that word semantics and overall sentiment of presidential speeches follow the trend of ongoing U.S. events throughout history. Specific keywords in the word embedding models highlight important historical events such as global conflicts, changing American policy, or differing political sentiment.

We attempted to split the speech data for the word embedding models by political party to see if there were any differences in similar terms by keyword according to ideology. However, we did not see any interesting results or glaring differences in word embeddings across parties, so we did not include these findings as a part of our results. We believe the reason for this is that only one party is represented per presidential term, so changing word semantics are already captured in the initial 20 year data splits we completed most of our word embedding analysis on.

When creating the visualizations, we initially used the raw sentiment scores for each speech. With this, we found that the plots were difficult to interpret as one speech's sentiment could be drastically different from the speech before or after it. As a solution, we created a new column in the dataset to hold the running average sentiment score so that we'd still be able to see the differences in sentiment between speeches, however the peaks and valleys on the chart would be less extreme.

In the future, we would like to do a more granular analysis of each speech to further investigate the specific events that cause the overall sentiment to decrease. This analysis could provide further insights on which specific events were viewed as the most serious in the country's history. Additionally, to validate the results of our analyses, we would like to investigate utilizing additional datasets containing information about the current state of the country. With this, we could cross check the true state of the country with the sentiment analysis of presidential speeches to assess how closely official statements from the President reflect the true state of the country.

Link to Data and Code Folder:
https://drive.google.com/drive/folders/1HGevaKGOp3JdVsPKf9-tcFCCLN2T72is?ths=true

List of Sources Cited

https://aclanthology.org/P12-3020.pdf
https://umich.instructure.com/courses/579491/files/folder/Week%202%20Readings?preview=28890454
https://gagantalreja.medium.com/creating-word-embedding-using-word2vec-18904a01c812
https://www.pnas.org/doi/epdf/10.1073/pnas.2120510119
https://www.sciencedirect.com/science/article/pii/S2090447914000550

Works Cited

Card, D., Chang, S., Becker, C., Mendelsohn, J., Voigt, R., Boustan, L., Abramitzky, R., & Jurafsky, D. (2022). Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences*, *119*(31). https://doi.org/10.1073/pnas.2120510119

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16). https://doi.org/10.1073/pnas.1720347115

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time Twitter sentiment analysis of 2012 U.S. Presidential Election Cycle. In Proceedings of the ACL 2012 System Demonstrations (pp. 115-120). Association for Computational Linguistics. Jeju Island, Korea.