Presidential Debate Analysis Final Report

By Michael Wermert, Tres James, Jordan Neely, Isaac Stone, Winston Boyd

I. Objective

The objective for this project was to use digital archaeological techniques to analyze previous United States presidential campaign debate transcripts. The purpose of the analysis was to determine if any debate factors (use of certain words, referencing certain issues, personal attacks, etc.) correlated with the outcome of the following election. Additionally, we would like to examine these debates over time to find any interesting patterns.

II. Discussion of Data Sources

We retrieved all of our transcripts for the presidential debates from the site https://debates.org/voter-education/debate-transcripts/. This website provided us with the transcripts for every televised presidential debate in American history. We also needed to retrieve the election results from https://www.archives.gov/electoral-college/1960. This website gave us the electoral college results for all of the presidential elections that we examined. We also used https://en.wikipedia.org/wiki/List_of_United_States_presidential_election_results_by_state to find the results of each election and data on the electoral college outcome. To analyze the transcripts, we wrote python programs that parsed the URL for each transcript and kept a word count between each of the participating candidates. We made use of some helpful python libraries to implement these programs and visualize the results, such as BeautifulSoup and matplotlib (as well as the NLTK toolkit).

We ultimately set out to perform several different analyses, such as:

1. How did the winning candidates' word choice differ from the losing candidates?
2. Did candidates who won the election talk about particular issues more than their opponents?
3. Are candidates who talk about certain issues more likely to win key battleground states?
4. Does Ad Hominem or Personal Attacks increase a candidate's chances of winning?
5. Under what circumstances are politicians more likely to use positive, neutral,and negative language?
6. Do candidates that use longer or shorter words on average in debates perform better or worse in elections?

III. Models and Algorithms

The first analysis that we performed was a word count analysis using an expanded version of the code that was used during Mini Project 1. This code obtained the word counts for each candidate in the debates, while word counts for moderators were filtered out. These results were displayed using bar charts from Mini Project 1. This can be used to determine which issues are talked about at a debate or at many debates throughout time.

The next analysis that was performed was a sentiment analysis using NLTK's vader toolkit. This toolkit is a rule based classification system that assigns positive or negative connotation to words based on ratings provided by real people. When a candidate's speech for all debates is plugged into this function, the percentage of positive, neutral,and negative language is displayed. This will be used to predict how much positive and negative speech appears in certain debates.

The next analysis that we performed was an analysis of the average length of a candidate's words. After we used the text from each candidate for the sentiment analysis, we iterated through this list of words and calculated the average word length of a certain candidate during his or her debates.

The final analysis performed on the debates was using a recurrent neural network (RNN). A recurrent neural network is a special type of neural network in that it contains particular perceptrons that are able to maintain a 'memory' of the data passed into it. Data is passed in to the RNN sequentially, making this model an appropriate fit for processing language data, which must be read or spoken sequentially in order to draw meaningful conclusions.

For the RNN model in this project, each input takes the first 100 words of a string of words, such as a sentence in the debate in this case. Each sentence is tokenized and each word within the sentence is mapped to an integer so that it can be fed into the network. Inputs with size less than 100 are padded with zeros. The output of the network is a floating point value between 0 and 1. If the value is less than .5, the sentiment output is negative. If greater than .5, the sentiment is said to be positive.

The model was trained on a Kaggle dataset consisting of 49,582 IMDB movie reviews with binary labels (positive sentiment or negative sentiment).

IV. Results

- **Word Count**

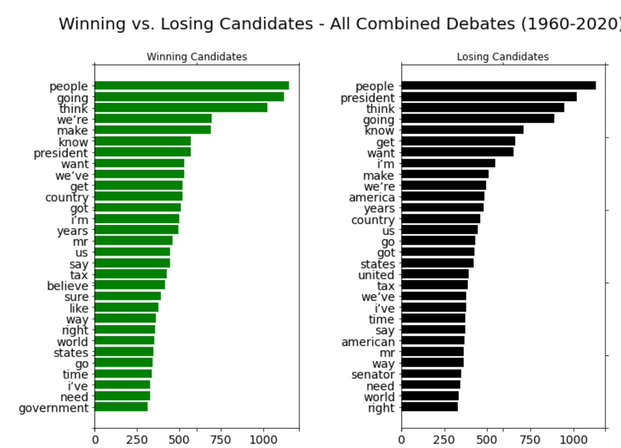


Figure 1. Word Count of winning vs. Losing candidates.

The results of our word count analysis showed that winning and losing candidates used very similar words throughout the debates, and word choice seems to have very little correlation to who wins the election.

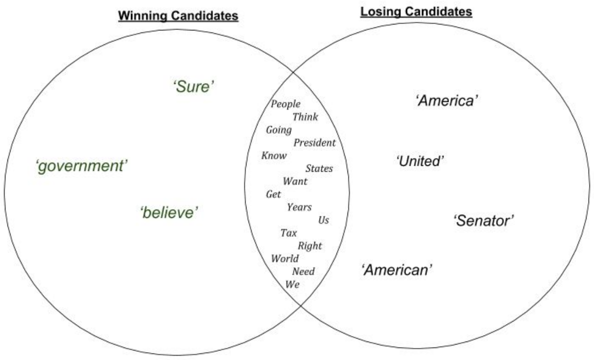


Figure 2. Venn Diagram of shared/distinct top words between winning and Losing candidates.

Figure 2 breaks down the top words that the losing and winning candidates had in common and those that were distinct for each side. This makes it more abundantly clear that the difference of word choice between the winners and losers is miniscule and inconclusive. The distinct words fail to prove any sort of recognizable pattern.

All of the candidates' most commonly used terms are fairly predictable generic ones, such as 'people', 'going', 'want', etc. This, if anything, proves that vague language seems to be the standard for all presedential debates (most of the individual debate charts had similar looking top words), though it also proves that using vague language doesn't necessarily give you the edge over your opponent in any way. Overall, word choice and count does not seem to be a very viable way to predict the outcome of an election.

When these results are broken down by party for all of the examined elections, the following word counts were calculated.

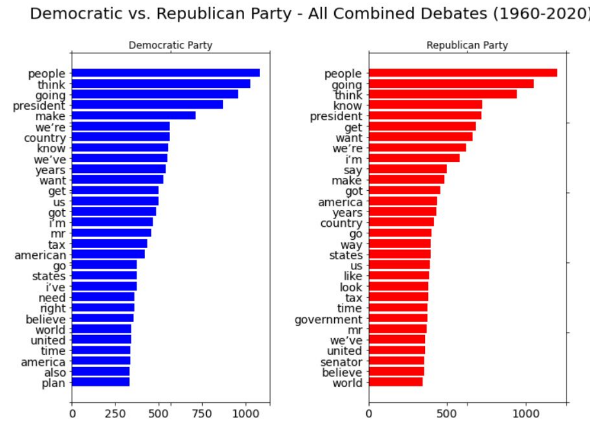


Figure 3. Word Counts By Party

Much like when the results are broken down by winning and losing candidates, the most commonly used words for both parties are very similar. The exclusive words for the Democrats are "american", "i've", "need", "right", and "plan". The exclusive words for the Republicans are "say", "way", "look", "like", "government", and "senator". The Democrat words are more interesting than the Republican words. Words like "right", "need", "americans", and "plan" could emphasize the Democrats' higher focus on working class citizens.

- **Sentiment Analysis - Vader**

Overall, winning candidates had a language breakdown of 8.6% negative, 77.3% neutral, and 14.2% positive. Losing candidates had a language breakdown of 8.3% negative, 77.4% neutral, and 14.0% positive. Overall, the results for each set of candidates were very similar. An overwhelming majority of the talk in these debates is considered neutral, which could be expected due to most of the talking being about the issues, while emotional statements that could be considered positive or negative are typically short. The results do show that winners of debates do tend to be about .3% more negative than losing candidates. This could indicate that being slightly more negative in a debate may help you with an election, but it would be a very small difference. It appears that sentiment analysis is not a great predictor of who will win an election.

One interesting pattern that this analysis showed is that typically, a challenging candidate will use more negative language than an incumbent candidate. This would make sense because a challenging candidate would want to attack the incumbent's record, which would lead to a higher amount of negative speech. Conversely, an incumbent candidate tends to use more positive language than their challenger opponent. This also makes sense because the incumbent candidate will often want to make his record look positive, so he will use more positive language. Overall, an incumbent candidate will be more likely to use positive language while a challenger will be more likely to use negative language.

- **Word Length**

An analysis of the average length of words for each of the candidates was performed as well. A winning candidate had an average word length of 4.3472 and a losing candidate had an average word length of 4.3546. These values are so close that I do not believe that word length would be a good indicator of whether or not a candidate will win the election.

However, there is an interesting pattern that was found when examining the average word length of candidates during debates over time. From 1960 until 2020, the average word length of candidates during presidential debates has decreased.

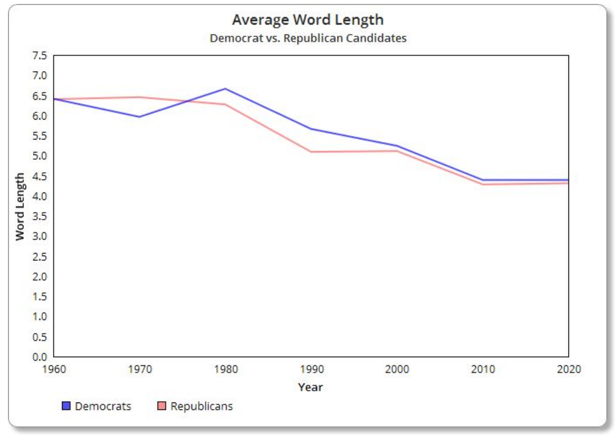
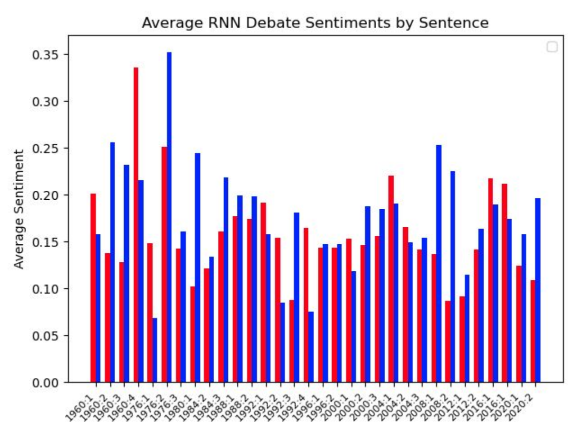


Figure 3. Average Word Length Over Time

- **Recurrent Neural Network (RNN)**

Each debate is divided into two lists, one for each speaker. Each list contains all the sentences spoken by one candidate during a particular debate. The words for each sentence in each list are tokenized, mapped to integer values, and fed into the network. Values between 0 and 1 (0 for negative sentiment and 1 for positive sentiment) are recorded. In the graph below, the average sentiment of each candidate for each debate by sentence is shown below in Figure 4.



As can be seen by the graph, the RNN interprets most of the debate rhetoric to be relatively negative. It is worth noting that there is not a single mean sentiment that is calculated to be positive, as there are no values above .5. This could be for several reasons. First, we consider the movie review dataset on which the RNN was trained. The language of a presidential debate and an online movie review differ by quite a bit. Because of the way language must be encoded before it is processed by the network, the network may not generalize well with words it has not seen and thus may provide an inaccurate classification. Another consideration is related to language but in a frame of context. The positive movie reviews contained language like 'great' and 'awesome' and other words associated with positivity that may not be considered positive in the context of a United States presidential debate.

V. Issues Encountered

Given the nature of our project, it became difficult to find what data we will train our models with. Our debates occurred from 1960 to 2020. Naturally, commonly used language has greatly changed over this period of time, so it was difficult to find a model that would be able to respond to all of the different words that have been used throughout all of the debates.

Another issue that was found was the lack of opinion polling data for the older elections. With the newer elections, we were able to find a decent number of opinion polls from both before and after the debates. However, for the older elections, this data was much less comprehensive.

VI. Future Work

In the future, it would likely be helpful to perform this analysis on other sources of speech that a candidate gives, such as speeches at campaign rallies, media appearances, and other long form speeches. It is clear that while presidential debates can often be important, there are other parts of a presidential campaign that are as or more important. If one were to perform similar analyses on other speeches, we might have more defined characteristics that we can use to predict winners.

Another interesting analysis of presidential debates could be an analysis of posture and facial expressions during the debates. While the speech during the debates is very important, facial expressions and body language can often play an important role in swaying peoples' opinions on a candidate. Libraries such as Keras can perform this kind of analysis. We could also use sound analysis libraries in order to determine if certain types of vocal tones affect a candidate's performance in elections.

While we only performed this analysis on presidential debates, we could also repeat this process for vice presidential debates and presidential primary debates. Primary debates may be an especially interesting case because they seem to affect candidates' polling numbers more than the presidential debates. Additionally, you could apply the techniques used in this study to debates that are not centered around politics, such as debates about science, art, and philosophy.

VII. Organization

Timeline

For the timeline, we mostly followed the plan we had set out at the beginning of the project with a few adjustments.

Sprint 1

- Set up guidelines for Analysis - i.e. charts, graphs, information want to be gathered, etc.
- Get data processing set up using python libraries

Sprint 2

- Have the complete analysis done by then Collectively collaborate piecing analyses together.
- Write up the progress report with team members.

Sprint 3

- Have finished the project's main purpose.
- Fix any bugs/mishaps we come across.
- Standardize the information to be processed in the final product.

Sprint 4

- Polish Analysis
- Extrapolate on the project
- Machine Learning Algorithms
- Finalizing the final report
- Complete whatever is left for the final presentation

Responsibilities/Contributions

Winston Boyd - Data/Analysis for 2012, 2016, 2020 Elections, Word Length/Count

Tres James - Data/Analysis for 1984, 1988, 1992 Elections, Electoral College Data

Jordan Neely - Data/Analysis for 1996, 2000 Elections

Isaac Stone - Data/Analysis for 2004 and 2008 Elections, RNN Sentiment Analysis

Michael Wermert - Project Manager, Vader Sentiment Analysis, Data/Analysis for 1960, 1976, 1980 Elections

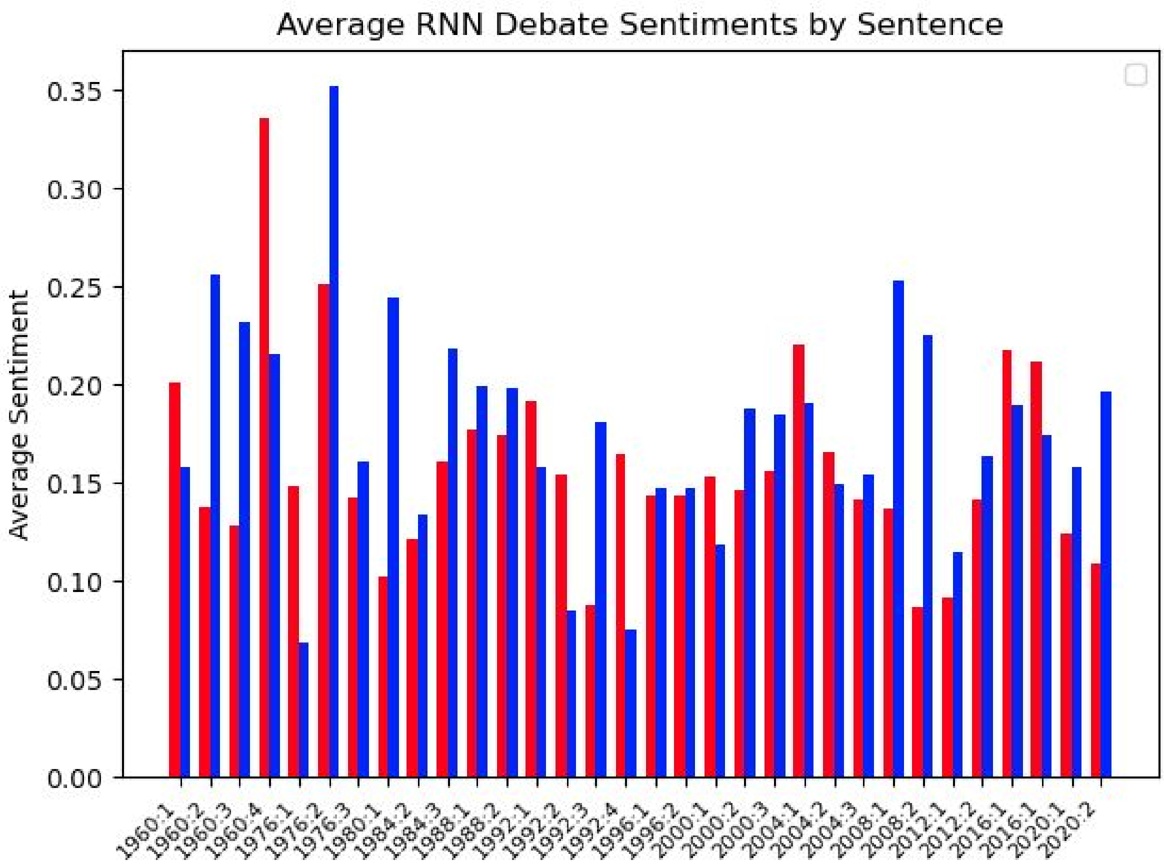


Figure 4. The labels along the x axis are to be interpreted as Year:Debate Number.