SENTIMENT OF THE UNION

by
Chase Rydeen

A thesis submitted to the faculty of The University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford
May 2018

Approved by

Advisor: Dr. Dawn Wilkins

Reader: Dr. Naeemul Hassan

Reader: Dr. Yixin Chen

ABSTRACT

As the machine learning and data science craze sweeps the nation, the implications and implementations are vast. This paper takes a look at both of them through the lens of a topic of national importance, at the very least for the United States. This topic is the words used by past Presidents of the United States, which are being pulled from their State of the Union Addresses. The focus of this research is on Natural Language Processing (NLP) and it's applied processes. Natural Language Processing allows for effective analysis of text-based data. Using NLP, a sentiment analysis was conducted on the Addresses to gain further insight into the tone used by Presidents over the course of history. This sentiment analysis ultimately resulted in a set of sentiment scores pertaining to major topics in the United States. These sentiment score sets were then input in to several different learning algorithms in an attempt to utilize Presidential Sentiment to predict political party affiliation. This paper shares the methodology used to conduct this sentiment analysis and discusses the tools created for the analysis and visualizations [Rydeen].

DEDICATION

To be completed.

ACKNOWLEDGEMENTS

Thank you so much to Dr. Wilkins for guiding me on this long journey and always being there when I need her. You are a true inspiration and have been so influential in my undergraduate career, and I'm so thankful for all your advice and help with this paper and this thesis.

Thank you Dr. Hassan for teaching me all I know about information visualization and D3.js to make the visuals for this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Language is our main form of communication and the words an individual chooses to express their ideas can be very telling about their mood and opinion towards a particular subject or issue. This is extremely pertinent for the President of the United States as his words and messages set the tone for the United States as a whole, and his words could possibly be analyzed to reveal deeper feelings about the matter at hand. This analysis has the potential to reveal underlying themes and patterns of speech in how presidents speak and what their word choice indicates about the State of the United States, as well as how they view certain issues and topics. The driving force behind the research conducted here was to see if there was a way to reliably predict a president's political party purely based on the words they use in State of the Union Addresses. This research can extend far beyond this central question as well, expanding to include more data sources to increase the accuracy of the predictions. On top of this, there is potential to predict further characteristics beyond political party such as ideology and other personality traits of the speaker.

The other crucial part of this research is more historical in nature in that each sentiment score derived for a president must be contextualized by the important historical events that occurred during their presidency. The biggest challenge here being separating out the relative importance of the president's own outlook on the world versus that of the event itself. Some presidents may have a more positive tone and outlook on the world and try to use their State of the Union address to encourage the public even if the events of the time may be dire, such as the Great Depression, so that was an interesting challenge to weigh the relative importance of each.

CHAPTER 2

CORPUS

2.1 State of the Union Addresses

The main textual data that was collected to be processed was all of the Presidential State of the Union addresses, from George Washington's first address to Barack Obama's last address. The text source was initially pulled from a Presidential Address Repository [Borevitz]. The text came in a large text file that contained every speech and it was split in to individual text files for each individual address to allow for easier processing. The addresses vary widely in length and content, which is also of significant note when analyzing and comparing these addresses across the timespan of the existence of the United States. George Washington's first address was just over a thousand words and seventeen paragraphs, whereas Barack Obama's final address was just over 5,400 words and was 78 paragraphs long.

2.1.1 Change in Purpose of State of the Union Address

The length is the most notable change in the State of the Union Addresses over time, but there are important factors to consider as well that could potentially impact how the addresses are given from year to year. When the Presidential Addresses first started with George Washington, it was not intended to be a recurring event [TETEN (2003)], but it soon began to grow to a tradition so that the president could publicly address the people and inform them of the current events of the country. Over the years, the Presidential Address has taken on many forms, in spoken word, in written letter, in radio broadcast, and, nowadays, on live television broadcast. The Presidential Address shifted from the yearly Presidential update, sometimes the only time people would hear directly from the President, to a formalized briefing to inform the public in an organized manner of the current state of affairs and push forward a President's agenda for the upcoming term [TETEN (2003)]. While that has always been a goal of the addresses, it has become more of the central focus over the course of time, due to technological innovations and changes in media coverage. Nowadays, citizens of the United States can read in real-time about the decisions of the Presidency and the Presidents political moves without needing to listen to an annual speech to become updated on their agenda and goals for the year to come. It is a subtle, yet interesting shift in how the addresses are approached, but even these purposes could change, depending on the person giving these orations, an important factor to consider also.

2.1.2 Presidential Personality

Another important factor in how the Presidential State of the Union Addresses are given is the personality of the President that is giving them. This is a rather intangible element of the speeches that can be hard to quantify but is very important to note. Most people have a certain disposition towards being more optimistic or pessimistic, and that can become apparent in the speeches given. The important topic being considered here is tone, which can be heavily influenced if the President giving the speech tends to be more realistic or optimistic in their outlook on the world. Some President's may see the State of the Union as a chance to rally the nation and project positivity and support for their platform for years to come, whereas others might see it as a good opportunity to have a nation-wide reality check and bring the citizens in-line with what needs to be done for the good of the nation [TETEN (2003)].

2.2 Statistical Summary

It is important to have an understanding of the speech data itself before diving in to this research, since otherwise it won't be as meaningful and it will be harder to draw conclusions. The full statistical summary for the data can be found in Table 2.1. This can be explored to search for trends in the data and familiarize oneself with an overall perspective on the data. Some information of note: There are a total of 230 Presidential Addresses given by 42 Presidents, making the average number of addresses per president 5. There are 26 Republicans and 16 Democrats, which makes their percentages 62% and 38%, respectively. The first three columns are self-explanatory and the latter two are described below.

2.2.1 Lexical Diversity

Lexical diversity is a metric that is used to represent the amount of unique words in any given passage of text and thus the overall complexity of the text [Johansson (2009)]. Lexical diversity is calculated by dividing the number of unique words in a text by the total length of the text. The resulting number is between 0 and 1 and the closer to 1, the more diverse the lexicon, so a value close to 1 can be interpreted as being more complicated to read. The patterns here can be confounding by sheer length of a text but it remains an important metric to see how complex a particular selection of text is. The nature of this calculation makes it more interesting when comparing two pieces of text that are similar in length to see the lexical diversity between the two. This calculation was performed for each State of the Union address and then all of the scores for each President were averaged together to get an average lexical diversity for each President.

2.2.2 Grade Level

Calculating grade level is a slightly more involved process that involves an algorithm that computes grade level based on two factors: average sentence length and average syllables per word. This formula was created in 1975 to determine the readability of documents for Navy enlisted personnel [Kincaid et al. (1975)] The first factor is relatively easy to calculate, but the second is slightly more tricky as syllables can be a lot more difficult to distinguish in plain text processing fashion. Luckily, there is a Python plugin called textstat with a built-in Flesch-Kincaid function that has a corpus of syllabled words and it was used to calculate grade level. You can see how the formula is used in Equation 2.1.

$$0.39\ (\frac{total\ words}{total\ sentences}) + 11.8\ (\frac{total\ syllables}{total\ words}) - 15.59 \tag{2.1}$$

2.3 Information Visualization

An important part of this research is also concerned with how best to display the resulting information in an effective and easy-to-understand manner. There is an entire field dedicated to how to best display technical information and data and how to convey it to large groups of people with little technical background [Fekete et al. (2008)]. This is important with data such as the sentiment score being processed here, as the long numbered sentiment scores are intimidating and without any context, data is meaningless. The context here is contained within the graph used to display the sentiment score data and interactive features were implemented to help users engage with the data in a more meaningful fashion. The data in this research is quantitative and since the Presidential Addresses are given in chronological order, time was used on the x-axis and the data lended itself nicely to a Scatter Plot. This scatter plot will be discussed more in-depth in the following section.

2.3.1 Word Cloud

A word cloud is a collage of words that displays word frequencies for a certain set of text data, with the relative size of each word being determined by the frequency with which that term is used in the text [Heimerl et al. (2014)]. An example can be seen in Figure 2.1 that shows the word cloud for all of Jimmy Carter's words he used for every one of his State of the Union addresses. A second example can be seen in Figure 2.2 where a term is selected and the word cloud dataset is restricted to the contents of that particular presidential address. Word Clouds are an interesting visual since they provide quick reference to see what a President's most used terms are, as well as being another interesting way to engage the data in a slightly different context. Word Clouds themselves are often criticized since it is a poor way to visualize data and it is hard to objectively compare two words in a word

President	# of Addrs.	Avg # Words	Lex Diversity	Grade Level
George Washington	8	2096.0	0.3762	18.55
John Adams	4	1801.0	0.369	17.925
Thomas Jefferson	8	2605.0	0.3376	18.0
James Madison	8	2729.0	0.3433	20.825
James Monroe	8	5326.0	0.2493	16.462
John Quincy Adams	4	7864.0	0.2327	19.25
Andrew Jackson	8	10708.0	0.2042	19.2
Martin van Buren	4	11411.0	0.2036	20.15
John Tyler	4	8560.0	0.2291	18.475
James Polk	4	18173.0	0.1525	17.275
Zachary Taylor	1	7678.0	0.2346	17.2
Millard Fillmore	3	10612.0	0.2224	16.967
Franklin Pierce	4	10545.0	0.2192	19.15
James Buchanan	4	14247.0	0.1797	15.05
Abraham Lincoln	4	6999.0	0.2639	13.675
Andrew Johnson	4	9690.0	0.2294	15.9
Ulysses S. Grant	8	8232.0	0.2391	15.938
Rutherford B. Hayes	4	8692.0	0.2363	16.325
Chester A. Arthur	4	5045.0	0.3252	13.6
Grover Cleveland	4	12478.0	0.2236	17.45
Benjamin Harrison	4	13881.0	0.1976	14.7
Grover Cleveland	4	14969.0	0.2121	16.35
William McKinley	4	16901.0	0.1977	15.8
Theodore Roosevelt	8	19793.0	0.1732	14.975
William H. Taft	4	17594.0	0.1868	17.025
Woodrow Wilson	8	4384.0	0.2768	15.05
Warren Harding	2	5738.0	0.2768	13.5
Calvin Coolidge	6	8707.0	0.2306	11.783
Herbert Hoover	4	6489.0	0.2566	14.15
Franklin D. Roosevelt	12	3991.0	0.3002	12.0
Harry S. Truman	8	8405.0	0.2321	10.475
Dwight D. Eisenhower	9	6103.0	0.2751	12.3
John F. Kennedy	3	5816.0	0.289	12.233
Lyndon B. Johnson	6	4917.0	0.2707	10.017
Richard Nixon	5	4002.0	0.2692	11.78
Gerald R. Ford	3	4649.0	0.2865	10.767
Jimmy Carter	4	11410.0	0.2427	11.05
Ronald Reagan	7	4731.0	0.2963	9.557
George H.W. Bush	4	4396.0	0.285	7.8
Bill Clinton	8	7528.0	0.2207	9.35
George W. Bush	9	4888.0	0.2883	9.122
Barack Obama	8	6738.0	0.2465	8.412
Donald Trump	1	5199.0	0.3043	8.4

Table 2.1: Presidential Summary Statistics

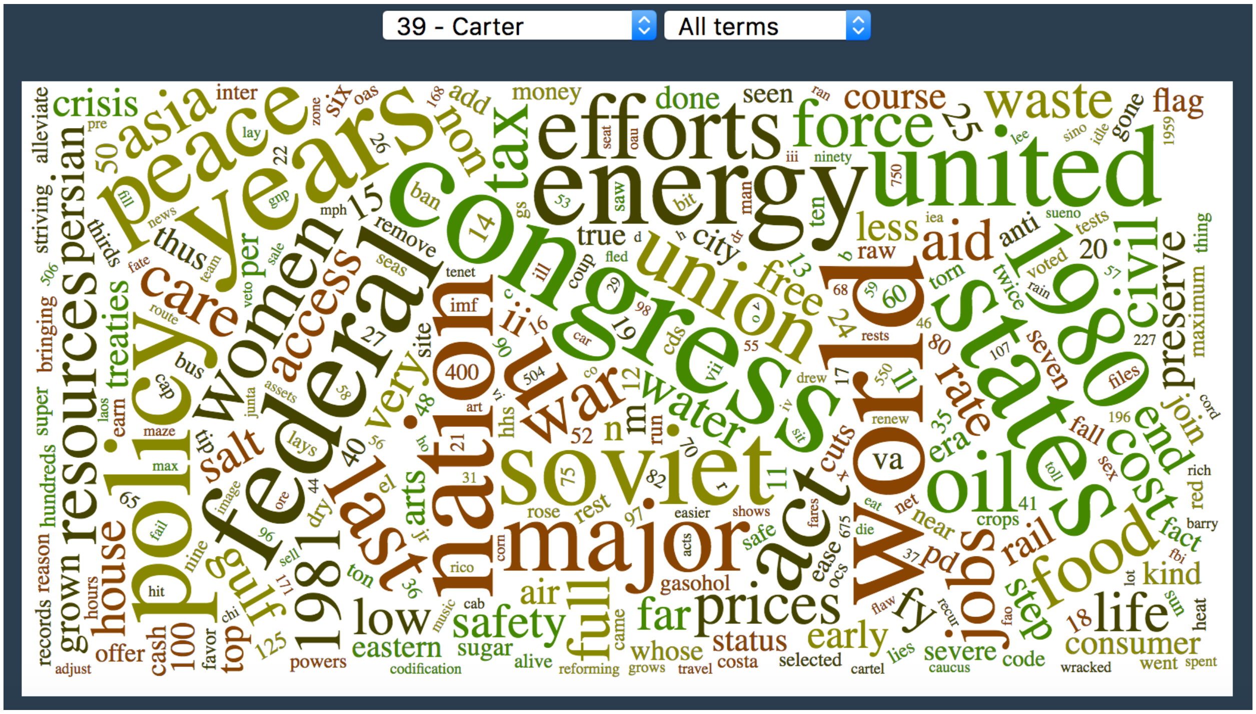

Figure 2.1: Example Word Cloud showing all terms for Jimmy Carter

cloud because the frequency values are encoded using area, which is a very difficult encoding for humans to interpret [Cui et al. (2010)]. In this case, the word cloud is used merely to complement the line plot visualization that will be introduced next chapter that provides insight into the main purpose of the research, and the word cloud provides a different way of visualizing the data source itself.

2.3.2 D3

D3.js (D3) is the JavaScript Library used to create the visualization mentioned in the previous section and another mentioned in a future section. D3 uses pre-built JavaScript functions to select, elements, create SVG elements, style them, or add dynamic effects or tooltips to them [Bostock et al. (2011)]. D3 also has a handy library that creates word clouds that was used in this research, it takes in an input array in JSON format with the words and their frequencies in decreasing order and draws the words with their relative sizes on the HTML canvas. There is a bit of a delay on the drawing of the word clouds, since instead

Figure 2.2: Example Word Cloud showing 1978 term for Jimmy Carter

of saved images of the word clouds, the program is actually drawing all of them in realtime and just swapping out the JSON data source depending on which President is selected in the dropdown menu at the top of the page.

2.4 Results

The bulk of this information is used later for processing, but it is important to understand the data source as well, before diving in to predictions using it. The importance and purpose of the State of the Union address is important here since it has a heavy-handed influence on the content and message behind the Presidential Addresses. It is important to note that the party breakdown, with 26 Republicans and 16 Democrats, makes their percentages 62% and 38%, respectively. These numbers aren't exactly correct, as the lesser-known and ephemeral early parties were placed in to either Democrat or Republican based on their policy positions. For example, Democratic-Republicans were assigned Republican as that party eventually became the common day Republican party, and the Whig Party was assigned to Republican as well as it was created from former of the Democratic-Republican Party. This was done in order to maximize the effectiveness of the prediction algorithm that will be discussed later on. The statistics here are important as they provide more insight into the data being processed and provide a more concise view into what is being handled. Table 2.1 has some intriguing patterns and trends to analyze and show.

2.4.1 Average Address Length

The average length of the Presidential Address has changed drastically over time, as its purpose and importance fluctuated. George Washington, when he gave his first address, didn't think that it would be a reoccurring event, but thought it necessary to inform the citizens of the current state of affairs of the country, and this precedent was followed for much of the early history of the United States [Freeman et al. (1948)]. The relatively short length of the early Presidential Addresses shows this, as it was short and brief. It was meant to inform the people of what is happening in the country and was used primarily to disperse information to the citizens. This slowly began to change over the years and the change can be seen in just the average number of words in the addresses. This change indicates the increasing importance of the State of the Union address as a chance to communicate with the citizens at large and use that attention to push an agenda and connect with the voters. The State of the Union address became a much larger deal as President's used it to communicate with the entirety of the nation to ensure them of the success of the nation and its status, peaking with Theodore Roosevelt averaging almost 20,000 words per Presidential Address. Shortly after, however, the length of the addresses had a tremendous drop-off from Taft at 17,000 words to Woodrow Wilson at 4,300 words, which can likely be attributed to the emergence of World War I. The country was involved in a major war effort and the fanfare and policy pushing of State of the Unions past were cleared out of the way for the focused messages of State of the Union addresses to come. These were defining times in the world, and with a major conflict to unite all people in the country, the State of the

Union addresses became more condensed and focused on the important aspects at hand. These shorter addresses were used to encourage the country and assure them of the success of the war effort and keep country-wide morale high and trying not to distract them for too long. From this major change and in to the modern era, the State of the Union address has stabilized around 5,000 to 10,000 words, keeping to an average length and the TV equivalent of roughly an hour to an hour and a half, long enough to keep people's attentions and effectively convey a president's reflections on the past year and goals for the next.

2.4.2 Lexical Diversity

Lexical Diversity, which was introduced previously, is also interesting to note here and it generally follows the same pattern as average length, just in the reverse fashion. As one would expect, the more words that are spoken, the less overall unique words are going to be spoken. This is most evident when examining the lexical diversity of George Washington and that of Teddy Roosevelt. George Washington had notoriously short State of the Union Addresses so his average lexical diversity was 0.3762, whereas Teddy Roosevelt has an average lexical diversity of 0.1732, which makes sense since his average length is almost ten times greater than that of George Washington's. This provides more important insight in addresses that are similar length to one another and provides a deeper insight into the speech-writing process and how word selection is important when communicating information to large swathes of people and needing to be considerate of their education levels.

2.4.3 Grade Level

Another metric that complements Lexical Diversity that needs to be considered is Grade Level, which was mentioned previously and it is computed using the Flesch Kincaid mentioned above. The scores seen here may seem rather high but it is understandable given the change in how Americans speak over time. Speakers in Early America were known to have a rather complicated way of talking and in order to make it to the office of President one had to be sufficiently educated to get elected by the public. This pattern shows in the high grade level throughout the early and mid history of the United States as most of the early presidents were college-educated, a rarity of the time, and had a more sophisticated vocabulary than the common man. Also the early speeches were often given in front of Congress and with no means of distributing the speech widely, the intended audience was mainly Congress, so the early Presidents did not really have a need to simplify their language to communicate effectively to the common man as was done by the newspapers that talked about the Address [Ziff (1991)]. The grade level gradually has decreased over time, which has as much to do with the greater reaches the address has, as it does with the way modern day media collects sound bites of presidential addresses. In the modern era, when a President gives a speech, only a small amount of the actual speech is rebroadcast when the media is discussing it, so the "sound bite" phenomenon has arisen in State of the Union addresses, which has had a transitive effect on the Flesch-Kincaid grade level calculation. The media only takes small snippets of what the President says to convey major policy positions, which has had the effect that most statements are kept short in order to summarize points and clearly convey what positions the President has in as short a form as possible [Paletz and Vinegar (1977)]. And since one of the calculations for the grade level calculation is average words per sentence, this brings down the grade level of the speech as the President attempts to become more clear in their purpose and position to effectively convey their thoughts and feelings in a short sound bite that could be taken from their speech.

CHAPTER 3

SENTIMENT ANALYSIS

The cornerstone for this thesis is sentiment analysis and this topic requires the definition of several terms and concepts that are important to the research at hand. Sentiment Analysis in theory sounds rather simple, process text and pull out the meaning based on the content of what was processed, but there are many intricacies that need to be addressed to fully understand the entire process [Liu (2012)].

3.1 Definition

Sentiment Analysis refers to the use of natural language processing and text analysis to systematically identify, extract, quantify, and study affective states and subjective information [Liu (2012)]. Sentiment Analysis has increased in popularity in recent years and is popular to use to review large sets of review / survey data to abstract major topics of conversation and controversy online. It can be an effective tool in summarizing a population's opinions and feelings towards certain issues and drawing conclusions from them. A basic task of sentiment analysis that can be leveraged into more complex tasks is determining the polarity of a sample of text data and classify it as positive, negative, or neutral [Wilson et al. (2005b)]. The process behind sentiment analysis is important and can be complicated depending on how in depth it is and how large of a dataset is involved, and before the process is addressed, there are certain principles and topics involved that need to be covered first.

3.2 Natural Language Processing

Natural Language Processing (NLP) is an important concept that is used heavily in sentiment analysis. NLP is primarily concerned with the interactions between human beings and computers and specifically how they process and discern the meaning of human language [Liddy (2001)]. Natural language data is abundant in our world today in the Age of the Internet and the vastness of it makes NLP extremely important to implement effectively to aid in understanding this large dataset. NLP takes on the difficult task of processing large text data and attempting to quantify the text data in different ways.

3.2.1 Natural Language Toolkit

NLP is the concept and the implementation in this project is the Natural Language Toolkit (NLTK), which is a Python library that offers NLP methods to process text and extract meaningful trends and patterns from the text of interest. The NLTK is implemented using Python, which is a simple, yet powerful language with excellent functionality for processing linguistic data [Bird and Loper (2004)]. Much of the meaning is derived using the NLTK but much of the processing is conducted in purely Python using lists of words to process meaning and sentiment. The task of processing text data comes with a few obstacles that can either obstruct meaning or complicate the processing by changing the meaning of words or phrases based on the context in which they reside.

3.3 Complications with Text Data

Text Data can be especially difficult to deal with and cause a lot of unforeseen issues when it is being processed. The inherent subjectivity of human language and speech is one of the largest obstacles that must be addressed when dealing with any text-based data. Much of the communication between human beings is subjective and it is up to the interpretation of the speaker and listener what the message is and they can have conflicting ideas on the meaning of some terms [Aggarwal and Zhai (2012)]. This potential for miscommunication is mirrored in NLP in that the determined meaning or value of some textual data could not represent the source it came from well, and there is not an objective reference to the weight of words to confirm the correctness of any one interpretation of the textual data.

Another important factor to consider when processing text data is sarcasm, which is extremely difficult to detect. Americans especially are known for their use of sarcasm and it can be sometimes impossible to parse such meaning out of text since the intonation is what indicates the sarcasm which is lost in purely text-based data. There are some subtle cues that can indicate sarcasm in text but it can really only consistently be caught if it is tagged as such [Riloff et al. (2013)].

A third and final complication that often arises is considering the context in which the text resides [Aggarwal and Zhai (2012)]. Context is everything when examining how people speak and trying to accurately access the thoughts and opinions of the speaker so it is important to take this context into consideration when performing a sentiment analysis, since it will affect any results that are obtained. Context here means the modifying words surrounding the word to be examined next in the analysis. Each word in a sample of text data can have any number of modifiers that can manipulate its meaning and fundamentally change what message it is conveying by adding certain words before or after the word. These modifiers can take on the form of intensifiers, such as “very”, that amplify the meaning of a word, or negators, such as “not” or “no,” that negate the meaning of a word.

3.4 Lexicon

There are several ways to conduct sentiment analysis, some of which do not require a lexicon but this research used a lexicon-based approach. An important tool that is necessary in conducting this analysis is a comprehensive lexicon. A lexicon is a database of words and accompanying features associated with each word [Taboada et al. (2011)]. These features associated with each word vary widely in what they indicate about the word, from part of speech to length to polarity score. The lexicon used in this research consists of a word and an associated sentiment score that is on the scale from -1.0 (negative) to 1.0 (positive), indicating how positive or negative the word is. These sentiment scores were compiled from several different lexicons and the results were constructed by surveying thousands of individuals and having them score a certain subset of words and combine those ratings into an average score for each word [Somasundaran and Wiebe (2010)].

3.5 Alternative Sentiment Analysis Approaches

There are other ways of conducting sentiment analysis without the use of a lexicon that can also be useful for conducting the analysis. The main alternative method is a comparative approach that compares each block of text data to one another and instead of giving them an objective score, ranks them according to a subset of rules that determine their ranking relative to the other samples of text [Wilson et al. (2005a)]. This approach puts much more focus on the context of what is being said and uses context to determine the polarity of language. This is the main approach used by many political science researchers, since it is far easier to compare politician's values relative to one another than use an objective dictionary to determine their stance on an issue. [Laver et al. (2003)]

3.6 Process

The process behind sentiment analysis, especially when it is lexicon-based, is rather simple to understand but there are a lot of hidden factors that must be considered. The most important and most influential part of this entire process is the lexicon, which was covered in the previous section. The lexicon is used as the basis for all the sentiment scores that are assigned in the analysis, thus its integrity and accuracy is central to the success of the analysis.

The analysis is started by inputting text data into the program. In this case, that text data was the Presidential State of the Union Addresses. The State of the Union is then separated into individual words, or "tokens" and all of those words are stored in an array for each sentence, and then each sentence array of words is stored in an address array. So, now there is an array of all of the sentences in the address, a for loop is used to iterate each of these sentences and search the lexicon for the presence of the words in each of the sentences. To begin, each word in the address is counted and added to a dictionary that contains all of the words and their respective counts in the address. While these words are being counted, the previous word in the sentence is cross checked against a list of negators and then intensifiers to see if either were used to alter the meaning of the current word. If a negator is found then, then an element is added to a dictionary that uses the word as the key and the value is the amount of times a negator is found before that word, and the same process is done for the intensifiers. To account for these intensifiers and negators, the word counts are decreased by one for each negator that was found before the word (cancelling out one use of the term), and increased by one for each intensifier (doubling the impact of the term). Now, using this dictionary of word counts, each word in the dictionary is compared with the lexicon to find its corresponding sentiment score. If the word is found, then the sentiment score associated with that specific word in the lexicon is multiplied by the number of occurrences of that word from the address, and this score is added to an overall sentiment score for the entire address.

However, if the word is not found, it could mean the word isn't in the lexicon, or that the word is in a slightly different form than what is stored in the lexicon, possibly being a verb in past tense with an -ed at the end. This problem was solved by implementing a basic stemming function in the for loop where if a word isn't found, then a letter is taken off the word and it is compared to all of the words in the lexicon once again. If this new word isn't found, then two letters are taken off and it is compared, and so on. This stemming only checks by removing up to three characters at the end of a word, covering most changes in ending. If the word is found after stemming then its sentiment score is multiplied by the count for that term and it is added to the overall sentiment score.

The entirety of the algorithm can be seen in Algorithm 1. This algorithm will be revisited later with some additions to account for topic classification.

Input: All State of the Union Addresses
Output: The sentiment score for each Presidential Address for each category.

```
open all .txt files and store them in lists of special category trigger words
for each address in the State of the Union Addresses do
    format address
    split address in to sentences
    for each word in the sentence do
        create word count for each word and store it in a dictionary
        if previous word negator then
            increment negator counter for that word by one
        end
        if previous word intensifier then
            increment intensifier counter for that word by one
        end
    end
    for each word in the dictionary do
        if word is in lexicon then
            if length of negators[word] != 0 then
                Subtract length from total count for that word
            end
        end
        if length of intensifiers[word] != 0 then
            Add length to the total count for that word
        end
        Calculate the Sentiment Score by multiplying the number of occurrences of the
          term by the score in the lexicon.
    end
end
```

Algorithm 1: Sentiment Analysis Algorithm

3.7 Scatter Plot

Scatter plots are effective at showing data over time and it allows for users to see overall trends in tone and compare the scores across presidencies to see tonal shifts over a president's tenure or how two presidents compared to one another. The data set lends itself to this representation and the result is a nice longitudinal summary of presidential tones over the course of history and you can see this scatter plot in Figure 3.1. The data being displayed isn't objective and it must be taken with a grain of salt because sentiment analysis is far from an exact science and the lexicon is objective but also doesn't take in to account

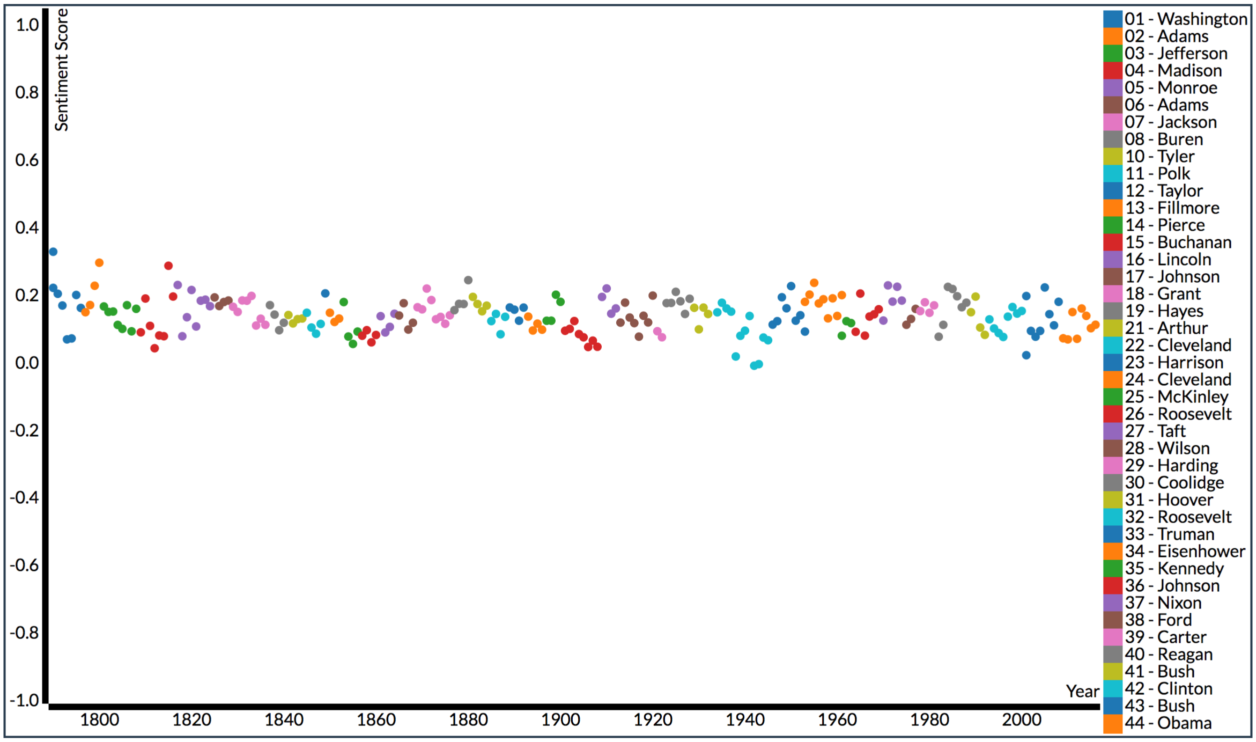

Figure 3.1: Scatter Plot

the change of in word usage and vernacular over time. The time period for these changes is a relatively short period of time in the context of language so the differences shouldn't be greatly significant in the shifting of tone but it is something to note. The scatter plot itself also allows for interaction in that the user can hover over a point and get detailed information about it, such as the President's name, the term and year that address was delivered, as well as the exact sentiment score. Figure 3.2 shows the hover feature over a specific Presidential address and showing the detailed information of the president's name, number, party, and sentiment score. Figure 3.3 shows the other hover feature wherein if a user hovers over a President's name then it shows all of their addresses and fades out the rest. Figure 3.4 shows a different scatter plot with the same data but color coded based on a president's political party to show different trends that follow political party lines.

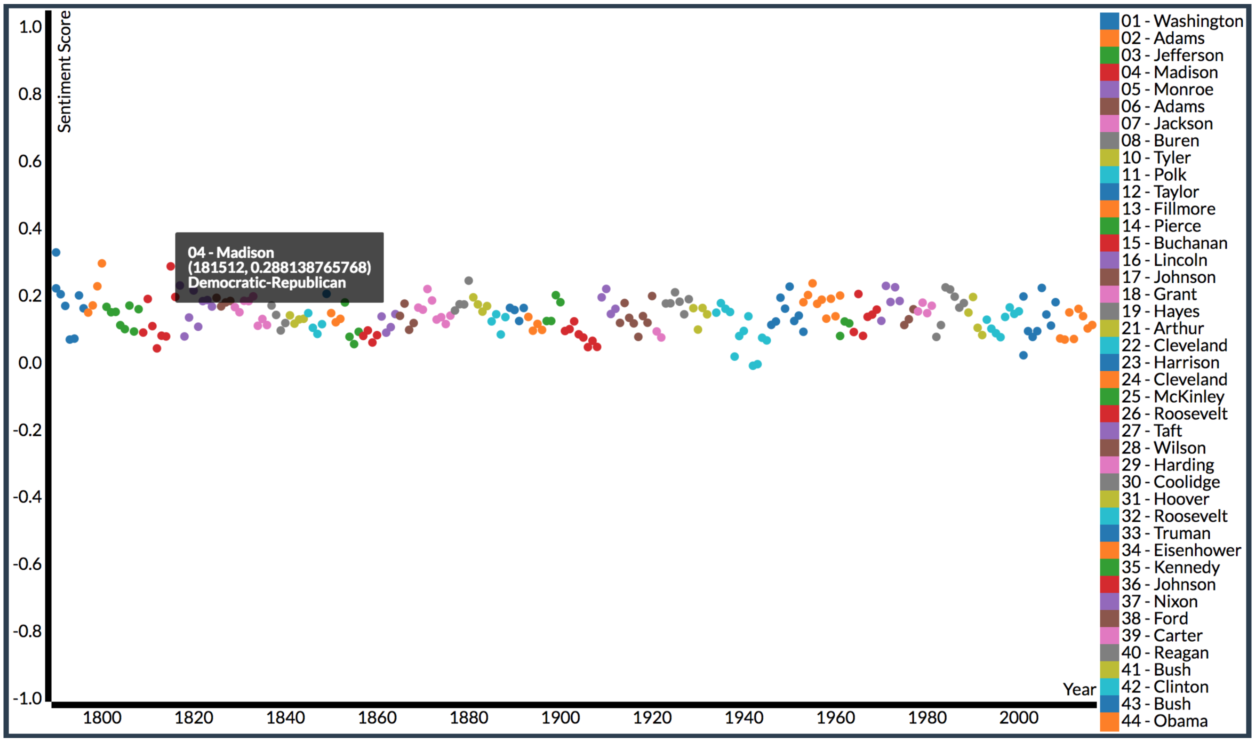

Figure 3.2: Scatter Plot (Hover over Address)

Figure 3.3: Scatter Plot (Hover over President Name)

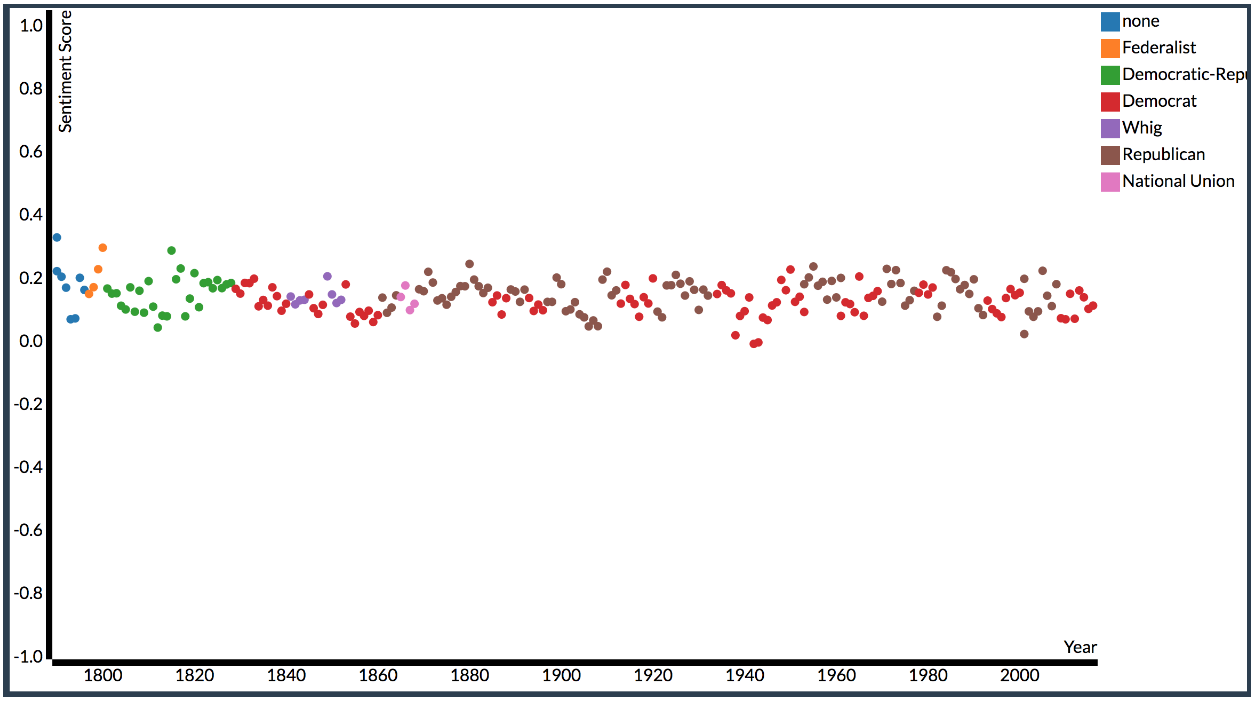

Figure 3.4: Scatter Plot Party Comparison

3.8 D3

D3.js (D3) is the JavaScript Library used to create the two visualizations mentioned in the previous sections. This tooltip functionality was used in the scatter plot demonstration mentioned previously to give extra information on each data point without cluttering the visualization itself. Another feature implemented using D3 is that if the user hovers over a President's name, then just that President's data points will be highlighted and the rest are faded out of the screen.

3.9 Results

This sentiment analysis is important to consider when analyzing the addresses and produces some interesting trends to dissect and investigate.

3.9.1 Overall

Generally, the trend for all addresses is that they hover around 0.00 to 0.30 sentiment score, which is telling of the general approach and purpose of the State of the Union Address. The address is used to convey the problems and issues that are impacting the country, and also to inspire hope that the country is making progress and that it is thriving. This slightly positive average overall describes an overall positivity among the presidents in their address to the United States citizens to inspire hope and confidence while also tending to the negative aspects of their presidency to show that change will be made.

3.9.2 Presidential Tone Trends

There are some interesting trends to examine in the tone of each president over time during their presidency. The overwhelming trend tends to be a generally positive tone for the early addresses that gives way to a more negative tone as the stark realities of the presidency become more apparent and the frustrations of being President start to show their effect in their speeches. This trend does have some notable exceptions, with John Adams, Rutherford Hayes, and George Bush starting lower and expressing more positive sentiment throughout their Presidential addresses. George Bush's evolution is most likely caused by 9/11 and the lingering effects that it had on his outlook on the presidency and the world at that time, but he got generally more positive over the course of his addresses following 9/11.

3.9.3 Historical Events

A large compounding factor that can have a heavy-handed influence on presidential addresses are major historical events such as the Civil War, the World Wars, the Great Depression, and 9/11. These events were dark and negative and it is interesting to examine the general tone of the Presidents during these times to see how they approached these grave and serious topics. They could be very realistic and inform the American people about the trials the country was facing, or paint them a more positive picture to encourage them in an attempt to boost the morale of the American people. The overall effect of positivity versus negativity could have many compounding factors that can't be fully investigated without a full psychological breakdown of each of the presidents. As such, the main consideration here is historical events since it is easier to examine their external effects on the United States and the Presidency than examining the preconceptions and biases felt by each President. These negative events are reflected in the addresses at the time, indicating the sensitivity of the Presidential outlook to major historical events. This isn't as true in the case of the Civil War as there aren't any major obvious effects in President Lincoln's speeches, likely due to the fact that it was Americans fighting Americans and his role was to attempt to unite the two sides into one cohesive unit so he attempted to avoid any accusatory and divisive language.

An interesting data point to draw attention to is that of FDR's two State of the Union addresses in 1942 and 1943, the only two Presidential Addresses to have a negative sentiment score in the entirety of the State of the Union's history, -0.01 and -0.005, respectively. This negativity shows signs of the hard times that was occurring during that time period in American History. FDR led the United States through the Great Depression, as well as World War II, and he won four presidential elections, so the people entrusted him and listened to his words. FDR had the responsibility to convey to the American people the importance and seriousness of World War II as it was happening thousands of miles away from the United States and many Americans felt it wouldn't impact them so they shouldn't send troops to die in a war they were not directly involved with [Roosevelt (1964)]. World War II was one of the most tragic events in world history and the tone of the President, and thus, the United States, reflects that.

Another interesting data point to pull out is that of George Bush's September 2001 Address, shortly after 9/11 had occurred. This was not a State of the Union address in the historical sense but it was needed given the events that had transpired so it is considered to be a State of the Union address [Bush (2001)]. The overall negative sentiment score of this address (0.02) when compared to his generally positive other addresses shows the dark and serious times that were apparent immediately after 9/11. With sorrow and remorse, George Bush addressed the people, and this language and purpose reflected itself in the tone of the words he chose to use in this address. This sadness, coupled with the calls for military action against the perpetrators of this tragedy tinged this address with negative and spiteful feelings, making it the third most negative address, and an extreme outlier in Bush's sentiment scores.

A third and final interesting point to examine is that of the very first State of the Union Address by George Washington in January of 1790. It is intriguing that the first State of the Union Address is actually the most positive of the entirety of the United States but from a sociological and historical perspective it does make sense. The United States is a young country and was in the state of being a democratic experiment since no one knew if it would actually work. It took George Washington, as the leader of this young nation, to inspire hope and foster confidence in this new country that it will succeed and democracy will prevail. An important role that he did not take lightly, as can be seen by the overwhelmingly positive approach he took to the first State of the Union address. Analyzing this in conjunction with his other sentiment scores is intriguing as they fall off sharply and he becomes generally more negative throughout his presidency, talking more of the stark realities of successfully building a strong country rather than the flowery patriotic speech of his first address. The trend does curl back up into generally more positive territory in his latter two addresses, an attempt to encourage the citizens and those who would assume his office in years to come with the confidence to drive America forward and make this democratic experiment a success.

CHAPTER 4

MACHINE LEARNING

Machine Learning has emerged as booming field in Computer Science that provides a lot of opportunities for innovation and growth. The important thing to know about machine learning is what is in the name: teaching machines to learn. Through various approaches and algorithms it is possible to feed these machines input data and coach them to predict outcome events without being explicitly programmed to do so [Hansen and Salamon (1990)]. Machine Learning algorithms come with a caveat though that unfortunately this research exhibits, and that is that effective machine learning is difficult because finding patterns is hard and often there isn't enough training data available to effectively train the algorithm to make predictions. The data here is large but rather minute compared to the large amounts of data normally used to train such learning algorithms. As such, the results achieved here aren't as strong as one would hope but this research establishes an approach that could be expanded and fed more data to achieve a more effective result.

Machine Learning can refer to many different topics as it is a broad field, but in this research the learning algorithms used were Neural Network, Naive Bayes Classifer, and Decision Tree. All of these fall into the supervised learning category of machine learning wherein a training set of data is input, along with the target outcome that allows the models to use this data and output target to learn how to predict the outcome [Dietterich (1998)]. Before the learning algorithms are discussed more in-depth it is important to understand the data that was produced to create the learning set and how it was used.

4.1 Topic Classifier Sets

An important part of the latter half of the preprocessing work for this research was the topic classifier sets. At first, the sentiment score was calculated for each presidential address with an overall score from -1 to 1, indicating their tone when delivering that address. After these were calculated, they were analyzed to look for trends in each president's tone to see if there were any interesting patterns. As an additional breakdown to see if there was any more context-specific information that could help determine a president's political party, topic categories were added to diversify the scores of the presidents.

Four Presidential Addresses were chosen (Washington, Lincoln, Kennedy, Obama) and manually read to discover what words were being used when talking about certain general topics within the United States. The twelve topics that were identified were: crime, economy, education, energy, environment, family, foreign affairs, government, job, religion, terrorism, and war. Text files were created using the trigger words that were collected for each major topic. The trigger words were pulled from the four addresses mentioned above and from various other addresses as they were skimmed through. During the algorithm, these text files are converted into arrays and as a sentence is being processed, it is scanned for these trigger words and if it has one of those words then it assigned to that topic. Then the sentiment analysis is conducted on each of the sentences within each of the categories to obtain a topic sentiment score for each President. The implementation of this part of the algorithm can be seen in Algorithm 2 on the next page, and the trigger words are used starting on line 7.

This processing was conducted on every address and the sentiment score for each topic was found for each President, which resulted in a vector for each address that had their overall sentiment score and the sentiment score for each topic covered in the address. These scores for each address were then averaged together to create an overall vector for each president that could be used for classification and learning to learn their political party. This vector consisted of 15 values (the President's name, the overall sentiment score, 12 of the topic sentiment scores mentioned above, and the President's political party) to be used for learning. For example, George Bush's vector was ['Bush', 0.1340853948691, 0.1302164475371004, 0.13079766376531318, 0.13296404930509467, 0.13296404930509467, 0.13743239837292998, 0.13949090904366207, 0.14036831555702362, 0.1434021113795079, 0.14341784426710505, 0.14306495549111 0.14278963481416337, 0.14030583896785492, 'Republican'].

Input: All State of the Union Addresses
Output: The sentiment score for each Presidential Address for each category.

```
open all .txt files and store them in lists of special category trigger words
for each address in the State of the Union Addresses do
    format address
    split address in to sentences
    for each sentence in the address do
        add sentence to 'overall' category
        if sentence contains category trigger word then
            add sentence to category
        end
        for each category do
            append list of sentences for that category to an overall list
        end
        for each topic in the overall list do
            for each word in the topic do
                create word count for each word and store it in a dictionary
                if previous word negator then
                    increment negator counter for that word by one
                end
                if previous word intensifier then
                    increment intensifier counter for that word by one
                end
            end
            for each word in the dictionary do
                if word is in lexicon then
                    if length of negators[word] != 0 then
                        Subtract length from total count for that word
                    end
                end
                if length of intensifiers[word] != 0 then
                    Raise length number of scores to the power of 2
                end
                Calculate the Sentiment Score by multiplying the number of
                  occurrences of the term by the score in the lexicon.
            end
        end
    end
end
```

Algorithm 2: Sentiment Analysis Algorithm

4.2 Normalization

As an added measure to clearly show the differences between different vectors, each of the values was normalized from -1 to 1 using a simple normalization algorithm. This normalization process aided in distinguishing the minute differences that manifest themselves when the data is more spread out on a greater range. The normalization method can be seen in Algorithm 3.

```
Input: Master array (an array of all the Presidential vectors)
Output: The Master array (Now with all values normalized)
for array in master do
    old min = min(array)
    old range = max(array) - old min
    new min = -1
    new range = 2
    array [float((n - old_min) / old range * new range + new min) for n in array]
    new_ master.append(array)
end
```

Algorithm 3: Normalization Algorithm

4.3 Neural Networks

Neural Networks take their name because they are trying to mimic and that is of a human's brain and its biological neural networks that allow it to make decisions [Hansen and Salamon (1990)]. This concept was mirrored and used to produce neural networks that are fed input data that is labeled as either exhibiting a behavior or not exhibiting a behavior and using that data to predict the unknown label of future inputs. The neural network has no inherent knowledge about the sentiment scores inserted into them, nor the political party label but it merely uses this data to learn patterns and uses these patterns to predict the political party of an unknown president using their sentiment scores. This first step of learning from data that is labeled is called the training phase. The training phase is important since the effectiveness of the algorithm relies entirely on the algorithm being trained correctly and effectively [Hepner et al. (1990)]. The goal is to have a diverse set of inputs to give the algorithm a range of data, and then tell it how many times to repeat

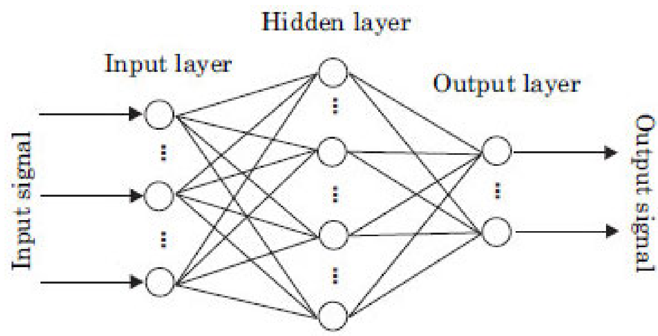

Figure 4.1: Neural Network Diagram

over the data to learn it. Finding the sweet spot of how many repetitions to utilize when having the algorithm learn the input data is very important, as too many repetitions causes the algorithm to confine itself to just the input data and it will lose the ability to generalize patterns to predict outcomes correctly, and too few repetitions prevents the algorithm from interacting with the data enough to draw meaningful patterns and conclusions from it.

A visualization for how a neural network works can be seen in Figure 4.1 [Juan et al. (2013)].

4.4 Naive Bayes Classifier

A Naive Bayes classifier functions in a very similar fashion to that of a neural network but it is less of a black box approach and more of a statistical approach. Using the input and target data, the Naive Bayes classifier uses a statistical model to predict values rather than strictly pattern recognition [Murphy (2006)]. Naive Bayes has actually been discovered to handle small amounts of data better than neural networks so it was important to add here since both have their strong suits in predicting values. Naive Bayes is a much simpler algorithm which can limit its performance and effectiveness as it attempts to fit its training data too closely, causing it to lose accuracy, whereas a neural network's complexity can actually overfit the data, which makes it weaker at predicting data outside the input data set.

4.5 Decision Tree Classifier

A Decision Tree Classifier functions in almost the exact same way that Naive Bayes does, but instead of predicting one output value, a decision tree examines the data to find steps it could take to make the correct prediction. Using these steps, a Decision Tree produces a list of steps it iterates through for each value and uses the outcome of each of the steps to predict the output value. This is effective with data that shows more trends and is sufficiently spread out, but this function struggled with this data as the decisions it made weren't clear and it overfit itself to the data which caused performance issues in this research [Dietterich (1995)]. A Decision Tree can be useful since it produces a model in a human readable fashion that gives insight into how it makes a prediction, which can allow for easier fine-tuning of the data and the algorithm to produce the best results. Much of machine learning can be a black box approach and this insight in to the inner workings of this algorithm simplifies it, but also limits it as this simplicity makes the algorithm not always as effective in its predictions.

4.6 Leave-one-out Cross-validation

In order to evaluate the effectiveness of the algorithm, leave-one-out cross-validation was used. This validation method works by iterating over the data and hiding one of the points of data and uses the remaining data points to predict the hidden one [Wong (2015)]. This is then repeated for each of the data points to be the hidden one. In this research, each address is represented as a vector of 13 numbers and two strings, indicating the sentiment scores for each category as well as the overall score and the final value is the political party the president belongs to. Then, using these vectors, one of them is hidden, and the rest of the vectors are used to predict the values for the hidden vector. This process is then repeated for each of the vectors until all of them have been the hidden one and had their output predicted. This validation method ensures the algorithm is working properly and can properly predict a set of values using the existing data set.

4.7 Results

The results from the machine learning algorithms were less than stellar but provide interesting insight into the problems at hand regardless of this. The breakdown of Democrat and Republican is 38% and 62% respectively. So, ideally the desired accuracy for an effective learning algorithm would be reasonably above 62% as you could successfully get 62% every time by predicting Republican for every single president. Unfortunately, the results achieved for these machine learning algorithms were 59.5% for the Neural Network and 35.71% for the Naive Bayes Classifier and Decision Tree. These accuracy numbers are less than satisfactory but there is much to say about the data being handled and how effective translating qualitative into quantitative data works. Text data at its heart is qualitative data since there is feeling and tone and intangible elements of speech that one can't quite quantify just yet but there is a way to do it. This research ran into many of these same roadblocks that come with translating text data into numeric data as some of this intangible meaning is lost and has to be reproduced mechanically to reach necessary conclusions about the data.

These results are less than astounding but it is interesting how much better the neural network performed than the statistical measure of the Naive Bayes. So the patterns drawn from the neural network were stronger indicators of party alliance and even though the data source was small, the neural network performed stronger even though typically the opposite is the case when comparing these two approaches as was mentioned previously. The Decision Tree Classifier has the same accuracy as the Naive Bayes as they both function similarly when the data set is small and they function very similarly in this research. Instead of creating a pattern to discern the vector values for each presidential party, the decision tree shows that it assigned the sentiment scores value to that category and if the values matched another one, it would look at the party of the matched one and assign it that, all the way down the list of categories. This overfitting caused the algorithm to focus too much on early results and not look at the whole data set before predicting a value which caused it to have an interestingly low prediction accuracy rate, worse than picking every party the same [Dietterich (1995)].

The concept and algorithms themselves are interesting but the accuracy and results are less than convincing about whether this can adequately be proven as a relation.

4.7.1 Vector Analysis

To add more context to the vector creation, here are two vectors from the calculations that will be interpreted. The two vectors can be seen in Table 4.1 below, and a full list of all of the vectors can found in the appendix [Here - To be completed].

The two excerpts are a small portion of the data collected but they both show interesting trends in tone across the different topics. Also, to be noted is the fact this these numbers are an average for all of the addresses given by the President and not single term. This is especially pertinent since George Bush had an extremely low sentiment score on his address immediately after 9/11 but his other ones were generally higher and averaged him out to a more positive sentiment score. Overall and across the categories, Abraham Lincoln has lower sentiment scores than Bush and actually his overall sentiment score is fairly higher than any one of his topic scores. This indicates a topic not featured here that he had an overwhelmingly positive tone on that averaged his overall score higher, or just a general positive mood not directed towards any specific topic.

Abraham Lincoln's presidency spanned the length of the Civil War so his tone during these State of the Union addresses can be viewed as the presidential perspective during this national outbreak of war. Lincoln does have a lower tone score than other presidents but it isn't as low as one might expect considering the events that were transpiring. This might indicate a generally positive tone as Lincoln tries to unify the country, but also follows the trend of Presidents having a generally more positive tone on their addresses overall. Also the tone on war is lower but also isn't out of line with any of the other categories really and this calls into question the exact calculations that go in to producing these scores. War can be a difficult topic to nail down in terms of meaning as most mentions of war play in to its destructive capabilities so trying to separate and distinguish positive and negative tone surrounding war can be difficult. This might play in to the numbers seen here and how they are generally close together since the trigger words and way the categories are sorted might need work to more effectively reflect the tone on specific topics.

Much of the same can be said for George Bush's vector with a few interesting differences. Bush's scores deviate below and above his sentiment score, with stronger positive sentiment occurring in Family and Religion, and lower scores in Economy and Government. George Bush was a Conservative and a large proponent of family values and was a devout Christian so the positive tone coming from those categories is not surprising. The lower scores for economy and war also make sense given the times as the economy was in a downturn at the latter half of his presidency and the Iraq war was going on during his presidency. This war sentiment score is deceiving, as was mentioned before, in how war is normally discussed, which should be considered.

It is interesting how the trend for war was equivalent for both presidents, being lower than the overall sentiment score. And George Bush had the reverse issue from Lincoln, where Bush's individual topic scores are all mostly greater than his overall score, possibly showing the hard negative pull of his address after 9/11. These vectors give interesting insight into the sentiment scores but also show some of the struggles that were encountered in creating these scores.

A more in-depth lexicon and a more comprehensive list of trigger words for each category would produce stronger sentiment scores that would be more effective in training a learning algorithm. A list of all of the trigger words can be found [HERE - to be completed] in the Appendix, as well as the lexicon used to calculate the sentiment scores can be found [Here - to be completed]. Another point of clarification that could improve the accuracy of the scores, as was mentioned above, would be to hone the polarity of objectively negative content involving war and other generally negative topics. In this research they were treated the same as other topics to keep the data consistent, but perhaps a more refined lexicon tailored to each topic could produce stronger results to make the predictions sought here.

Last Name	Lincoln	Bush
Overall	0.1207	0.1341
Government	0.1076	0.1302
Economy	0.1100	0.1308
War	0.1043	0.1330
Terrorism	0.1043	0.1330
Jobs	0.1043	0.1374
Education	0.1035	0.1395
Foreign Affairs	0.1031	0.1404
Environment	0.1078	0.1434
Energy	0.1078	0.1434
Family	0.1063	0.1430
Religion	0.1068	0.1428
Crime	0.1051	0.1403
Party	Republican	Republican

Table 4.1: Presidential Average Sentiment Score by Topic

CHAPTER 5

CONCLUSION

This research has been intriguing and interesting but has fallen victim to many shortcomings that come with textual data and human emotions. There just might be a clear correlation between a President's tone on a specific topic in the United States and their political party but the results found here cannot prove such a thing. The art of converting qualitative text data into actionable quantitative data is still a process in its infancy and many advancements are to come in this field before it flourishes into a more accurate and effective prediction method.

5.1 Complications

The complications arose mostly from the text data and manipulating it effectively to translate it into numbers while retaining as much meaning and context as possible. There is only so much meaning and interpretation that can be derived from just the text without consideration for the socio-political climate at the time that the speech was given, which is a much harder problem to solve and quantify. The potential for this research to aid in political science research on presidential profiles is high, but as a stand-alone method for interpreting Presidential party alignment it needs more work and fine-tuning to do that effectively.

Another major pitfall that this research ran into was not having a large enough data source to compile specific profiles for each president to form their political profiles in stronger ways to shape a political party position. The scope of the dataset was limited to State of the Union Addresses as they are consistently delivered each year by the president so the standards were understood and known for Presidents past and future. This consistency is important since the speeches can be interpreted and analyzed given the same basic list of information to look for in this address. Also the typical fashion in which the State of the Union is delivered mandates the President address each major topic of interest concerning the United States, thus lending itself to be analyzed in this automatic fashion. A possibly more effective, yet time-consuming approach, would be to include personal writings and other speeches given by the president and discern them for meaning and add them to the corpus of text data analyzed. Some of these documents would be short and some of them would need to be manually tagged for meaning depending on what the content of the speech was, but perhaps this would provide greater insight into the Presidential profile and thus create a stronger party profile on which to predict Presidential alignment.

Whether the shortcoming of these predictions come from a lack of data or a lack of correlation is impossible to tell and no such conclusion can be made at this time. Perhaps, their tone when speaking on certain topics can show their party alliance but no strong evidence has been found thus far. And there could never be an accurate gauge that is reliable enough to predict Presidential party given the nature of how a President gets elected in the first place. Most Presidents are moderate enough where they can swing at least a portion of the vote in their favor. So, although a president might have particularly strong feelings on some categories they have more moderate opinions on others that average out to a moderate take on many things. This inherently moderate nature of the President doesn't bode well for predicting their party alignments but with diversified text data this could potentially be rectified. This research could also be reproduced using the Supreme Court decisions as text data and predict party alliance based on how the Supreme Court Judges decide since their political alignment is better known and can be pinpointed more resolutely than the President's since they decide on every case and have more consistent output of text data to analyze.

5.2 Future Work

There is much that could be done to continue this research to make it more effective, and also to make it more interesting and intriguing. Some important future work would be adding additional visualizations that further breaks down all of the data discussed here. There is a great amount of it and expanding upon these visualizations would make it much more effective to look at and analyze. These visualizations would incorporate historical events and allow for tracking of a president's tone over time as it correlates to major historical events in the United States, as well as the world. These visualizations could also include a per year approach that allows a user to look at a particular year to see the president and sentiment score as well as other important economic and social information to examine the correlations between the well-being of the nation and the overall attitude of the President.

Another intriguing avenue to pursue would be to look at House and Senate majorities versus ruling presidency and how many bills were passed and how many laws were implemented, and compare that to the tone of the President. Perhaps to see if the frustrations of getting bills and laws rejected would reflect itself in a more negative tone of the President. An interesting expansion to this research that might warrant a whole new thesis itself would be exploring the Supreme Court Decisions and crafting party alignment using the Supreme Court Justices' decisions and statements and attempting to use that to predict a person's political party alignment based on the content of their speeches or writings.

5.3 Final Thoughts

This research has been intriguing and rewarding and provided quite a lot of obstacles and challenges. There is still much to explore in Natural Language Processing and Machine Learning as the surface was only scratched throughout this research. The overall question of Presidential tone and political party alignment still remains to be explored and hopefully this helps as a starting point for future research into this immensely interesting avenue of research.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aggarwal, C. C., and C. Zhai (2012), *Mining text data*, Springer Science & Business Media.

Bird, S., and E. Loper (2004), Nltk: The natural language toolkit, doi:10.3115/1219044.1219075.

Borevitz, B. (), Presidential address repository, `http://stateoftheunion.onetwothree.net/appendices.html#source`, accessed: 2017-09-30.

Bostock, M., V. Ogievetsky, and J. Heer (2011), D3 data-driven documents, *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2301–2309, doi:10.1109/TVCG.2011.185.

Bush, G. W. (2001), 9/11 address to the nation, *American Rhetoric*.

Cui, W., Y. Wu, S. Liu, F. Wei, M. X. Zhou, and H. Qu (2010), Context preserving dynamic word cloud visualization, in *Visualization Symposium (PacificVis), 2010 IEEE Pacific*, pp. 121–128, IEEE.

Dietterich, T. (1995), Overfitting and undercomputing in machine learning, *ACM Comput. Surv.*, 27(3), 326–327, doi:10.1145/212094.212114.

Dietterich, T. G. (1998), Approximate statistical tests for comparing supervised classification learning algorithms, *Neural computation*, 10(7), 1895–1923.

Fekete, J.-D., J. J. Van Wijk, J. T. Stasko, and C. North (2008), The value of information visualization, in *Information visualization*, pp. 1–18, Springer.

Freeman, D. S., J. A. Carroll, and M. W. Ashworth (1948), *George Washington, a biography*, Scribner.

Hansen, L. K., and P. Salamon (1990), Neural network ensembles, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10), 993–1001, doi:10.1109/34.58871.

Heimerl, F., S. Lohmann, S. Lange, and T. Ertl (2014), Word cloud explorer: Text analytics based on word clouds, in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pp. 1833–1842, IEEE.

Hepner, G., T. Logan, N. Ritter, and N. Bryant (1990), Artificial neural network classification using a minimal training set- comparison to conventional supervised classification, *Photogrammetric Engineering and Remote Sensing*, 56(4), 469–473.

Johansson, V. (2009), Lexical diversity and lexical density in speech and writing: a developmental perspective, *Working Papers in Linguistics*, 53, 61–79.

Juan, W., W. Pute, and Z. Xining (2013), Soil infiltration based on bp neural network and grey relational analysis, *Revista Brasileira de CiAtextordfemeninencia do Solo*, 37, 97 – 105.

Kincaid, J. P., R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom (1975), Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel, *Tech. rep.*, Naval Technical Training Command Millington TN Research Branch.

Laver, M., K. Benoit, and J. Garry (2003), Extracting policy positions from political texts using words as data, *American Political Science Review*, 97(2), 311–331.

Liddy, E. D. (2001), Natural language processing.

Liu, B. (2012), Sentiment analysis and opinion mining, *Synthesis lectures on human language technologies*, 5(1), 1–167.

Murphy, K. P. (2006), Naive bayes classifiers, *University of British Columbia*, 18.

Paletz, D. L., and R. J. Vinegar (1977), Presidents on television: The effects of instant analysis, *Public Opinion Quarterly*, 41(4), 488–497.

Riloff, E., A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang (2013), Sarcasm as contrast between a positive sentiment and negative situation, in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 704–714.

Roosevelt, F. D. (1964), *Four freedoms speech*, Project Gutenberg.

Rydeen, C. (), Sentiment of the union visualizations, `https://turing.cs.olemiss.edu/~dcrydeen/thesis/index.html`, accessed: 2017-10-15.

Somasundaran, S., and J. Wiebe (2010), Recognizing stances in ideological on-line debates, pp. 116–124.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll, and M. Stede (2011), Lexicon-based methods for sentiment analysis, *Computational linguistics*, 37(2), 267–307.

TETEN, R. L. (2003), Evolution of the modern rhetorical presidency: Presidential presentation and development of the state of the union address, *Presidential Studies Quarterly*, 33(2), 333–346, doi:10.1111/j.1741-5705.2003.tb00033.x.

Wilson, T., J. Wiebe, and P. Hoffmann (2005a), Recognizing contextual polarity in phrase-level sentiment analysis, in *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pp. 347–354, Association for Computational Linguistics, Stroudsburg, PA, USA, doi:10.3115/1220575.1220619.

Wilson, T., J. Wiebe, and P. Hoffmann (2005b), Recognizing contextual polarity in phrase-level sentiment analysis, in *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pp. 347–354, Association for Computational Linguistics.

Wong, T.-T. (2015), Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation, *Pattern Recogn.*, 48(9), 2839–2846, doi:10.1016/j.patcog.2015.03.009.

Ziff, L. (1991), *Writing in the New Nation: Prose, Print, and Politics in the Early United States*, Yale University Press.