Presidential Pandering?: An analysis of State of the Union addresses using machine learning techniques

Author: Jeff Bertucci

**Introduction**

Outlined in the United States Constitution, the president “shall from time to time give to the Congress Information of the State of the Union, and recommend to their Consideration such measures as he shall judge necessary and expedient”. This has been interpreted by many to mean that at least once per year, the president should make a formal address to the congress to outline their intentions for the coming year and guidelines on how they will be governing for their term. Originally, this was done in the form of an administrative report that was sent to congress on what departments were doing and the spending in relation to those departments. This process was changed when Woodrow Wilson delivered the first in-person state of the union address in order to rally support for his fourteen points, and the format is still used today. With changes in technology, this address was able to be distributed via radio, then television, and currently webcasts and streaming. More people than ever before have access to view their president as they update the masses.

With the change in audience, presidents have followed Wilson’s lead in engaging with the public, rather than keeping their decisions insulated, and now once per year, can make a stump speech where everyone in the country could be watching. This is the time they have to make their pitches to the American people; they can talk directly and succinctly about what they want to accomplish. They can also choose to rally their base, or publicly pressure their opponents. They can give insight on a divisive topic or they can share a story about their constituents. However the president’s team decides they want to play it, this can be the time where they have a direct platform to the American people and is their opportunity to sell themselves as a leader.

As speeches have gotten more technologically complicated, so too have the reporting methods on state of the union addresses. It is important for the news media to stay relevant when it comes to their reporting, and they must always stay one step ahead of the public knowledge in their analysis. Before the speeches were readily available to the public, the news media was used as an intermediary to deliver the official’s words to the public. This is no longer enough to keep the public engaged. Instead, news organizations now need a cast of analysts that are the best at what they do to deliver interpretations of everything from policy to body language to fashion choices. Running and engaging with this wide variety of analysts can be exhausting. When a news company can spend money to bring in an expert, it is a guarantee that expert will be called upon. As such, mountains of data can pile up from multiple directions in order to interpret a single one-hour speech. Given that, organizations need to keep the news cycle in mind because all this information will not be relevant after the cycle is over for the speech or topic. Provided that the news cycle for stories is slowly shrinking while the amount of data is quickly rising, it is important for organizations to come up with new innovative ways to parse the data while also keeping costs low.

Machine learning is in the position to assist with this issue. With a good model, it may be possible for analysts to put the speech through and receive to the minute analytical updates on everything the president is saying. This can be used in a variety of ways, from fact-checking, to cross-check a president’s words with earlier statements, and to compare with a vast corpus of other speeches of the same variety. These models can be used on the analysis side as well as on the generation side of the speech. With the correct model, proposed lines can be fed, and the potential impact can be calculated and measured against other lines of a similar nature. As such it is important to be making these steps in order to stay ahead of the game and make the most positive impact possible.

**Analysis**

**About the Data**

The data used consists of the text of the 4 most recent presidents’ first state of the union speech as officially posted by the White House. The mentality in choosing these specific speeches comes from the thought that when a president first begins their term, they will not be seeking re-election for another 4 years, meaning that they have the space to outline more of the policies they want to accomplish and less time campaigning. They also have a positive outlook on the future where they may believe they have more room to enact change before the reality of the position and the politics around policies come into play. One note to make on this fact is that these speeches are not technically considered to be state of the union addresses despite them being labeled as such. This is due to a precedent set that a freshman president does not make an official address, and instead it is more so considered to be an informal outline. They are still promoted as state of the unions and have the same impact as the speeches given at the same time the following three years.

The text data is first aggregated into a single aggregate document as well as saved individually in a corpus. In the aggregate, the sentences are separated line by line so each sentence is an entry. Two columns are then added. The first is which political party the president belongs to, and the second marks whether a statement is an action item or rhetoric. An action item denotes a direct promise the president makes or a specific thing that they have already done to accomplish their goals. Rhetoric, on the other hand, is a statement that is made to sway the public over to the president’s way of thinking. This can be a story, statistics, an appeal to values, or definitions. Essentially, the term of rhetoric is used in this case as a catch-all term to describe all statements that are not action items.

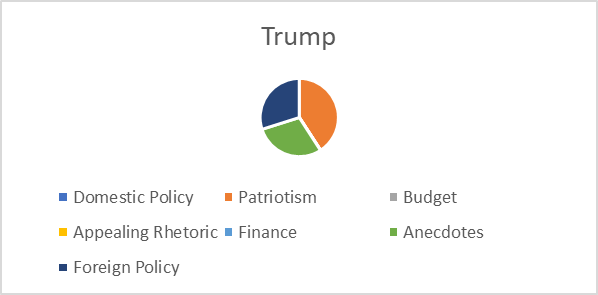
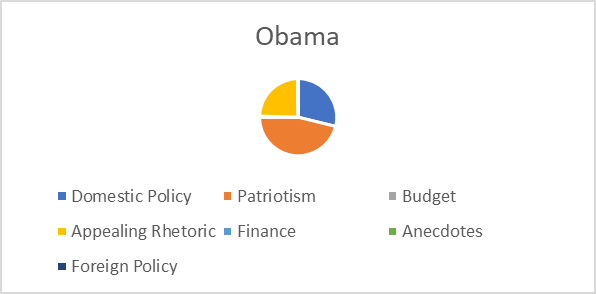
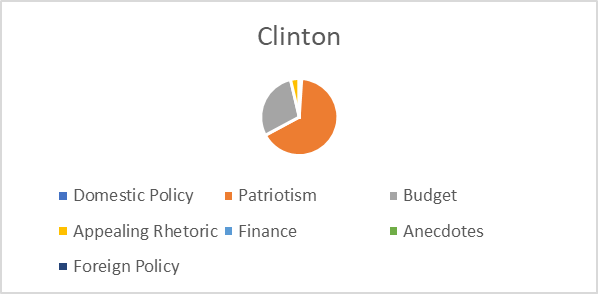
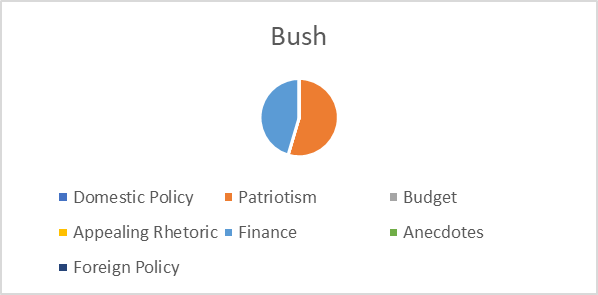
This column was filled out by hand by a single person interpreting the text data. In so doing, there are certain flaws that would need addressing. Specifically, the interpretation was done a single time by one person due to timeline constraints, meaning that there will not be a good metric to qualify this classification. As it was done by a single person, there is also the consideration to make of bias. While an attempt was made to fairly categorize each statements (they were scrambled so that the speaker was unknown) there is still the potential for lurking biases.

After the data to be used is set up, some basic information is taken regarding the statements. In total, there are 1235 sentences being analyzed. Of these sentences, republican presidents account for 43% of them and democrats account for the remaining 57%.

This means two things. First, democrats may be slightly overrepresented by the data due to Clinton and Obama’s propensity for long speeches, and as such they may have more action items to work with as shown below:

This disproportionality may mean that models used to predict action items may tend to better predict democratic actions better than it can predict republican actions if that is a factor. In a comparison of action items to rhetoric for the speeches of both parties, it was found that for both parties are within a few points of 25% of their total statements being action items and 75% of their speeches being rhetoric.

The text data saved into the corpus is converted into a mallet format in order to do topic modeling. Through multiple iterations and fine tuning of parameters, seven topics where separated out in order to better classify parts of a generic state of the union address and twenty words were used as markers in each topic in order to make that classification. The topics found were classified thusly: Domestic Policy, Patriotism, Budget, Value Rhetoric, Finance, Anecdotes, and Foreign Policy. To get a better idea of the data that will be worked with, each speech was classified in percentages based on these topics and the result is shown below:



While all four presidents have a hardline on patriotism in their speeches, the other things they focus their time on is up for debate. Obama in his speech, for example, decided to place more emphasis on domestic policy issues such as health care and pushing forward value rhetoric. In comparison, in Trump’s speech, his focuses where more outward, looking at foreign policy issues such as immigration while also telling stories about individual constituents. President Bush had a focus on taxes and business finances while Clinton focused mainly on the US budget. This shows that each president has different tactics they use when they make their pitches, meaning that classifying each based on the others has the potential of being a difficult task. Word clouds are shown below to highlight the similarities and differences in the speeches:

Clinton:

A close up of a logo

Description automatically generated

Bush:

A close up of a sign

Description automatically generated

Obama:

A close up of a sign

Description automatically generated

A picture containing text

Description automatically generatedTrump:

To prepare the classification process, the text, party and action data is loaded using python. The labels are adjusted for clarity and the text data is scrubbed of all punctuation and other symbols. A stemmer is not used in this process in order to maintain the semantic meaning of words that could be relevant to determine action words. Common English stop words are removed in this process in order to cut down on the processing and to not overrepresent common words within the decision-making process. Two vectorizers are employed to make a term matrix for these statements, the first uses unigrams while the second uses unigrams and bigrams. All other features remain the same. The data is then separated into training and testing groups of 70% training and 30% testing. The labels are removed from the testing data and the party data is removed from both sets as well. For testing, a multinomial Naïve Bayes model and an SVM model are used.

**Models**

**Naïve Bayes:**

**Model 1.1-Unigram Model 1.2-Bigram**

**Accuracy: 75% Accuracy: 67%**

**A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated**

**SVM:**

**Model 2.1-Unigram Model 2.2-Bigram**

**A screenshot of a cell phone

Description automatically generatedAccuracy: 77% Accuracy: 78%**

**A screenshot of a cell phone

Description automatically generated**

**Results:**

The best results stem from the SVM model that used both unigrams and bigrams. This is most likely due to the increase in number of variables the model gets to work with and cut out since the SVM model can choose which features to employ. In the case of the Naïve Bayes model, it would seem that the excess data was more of a burden than a help as accuracy from the unigram model dropped by 8% in moving from one model to the next. Ideally, these models would be able to predict at an accuracy significantly above 75% as that is the percentage of rhetoric sentences. Only the SVM models were able to accomplish this with the data, and not by much. However, none of these models chose to classify all statements as rhetoric meaning that a metric of significantly higher than 50% (or random chance) is a positive result although not ideal.

The SVM model does look to be the best choice in either decision, but his may be due to the restrictiveness of the data. Should more address be utilized and classified in this way, there is a chance that these models can improve substantially and the Naïve Bayes models will have more standardized data to work with.

Working with these models, the terms that had the most impact are listed below:

**Unigram:**

control: 2.056198740818161

hundreds: 2.056198740818161

improved: 2.056198740818161

initiative: 2.056198740818161

iraqi: 2.056198740818161

principles: 2.056198740818161

proposing: 2.056198740818161

way: 2.0381458214039405

including: 1.922667348193638

week: 1.922667348193638

**Uni/Bi-gram**

was: 2.4503313593480343

this is: 2.2271878080338254

just: 2.1530798358801038

what: 2.0730371282065674

always: 2.0730371282065665

big: 2.0730371282065665

interest: 2.0730371282065665

clean energy: 2.021307434015533

equal: 2.021307434015533

no longer: 2.021307434015533

Looking at these words, it can be reasoned which way they swayed the model. Words such as proposed, initiative, and week are examples of words that have immediacy. Ie: “We take an initiative to…” or “This week we will…” or “I propose we..” whereas terms like “principals”, “hundreds”, and “including” imply less of a direct action and more of an appeal to principals, statistics, or interest groups. These are similar terms to the uni/bigram approaches, but it would seem that for the uni/bi-gram approach, there is more of a focus on categorizing as rhetoric more than categorizing action items. This would explain why the bi-gram approach was better for the SVM model and worse for the multinomial naïve bayes model.

**Conclusion**

In the current political climate, it is important to be as knowledgeable as one can be when it comes to what one says. It is necessary to hold politicians accountable for the words they say, because it is their job to say those words, and that is what their constituents will hold them to. Should a someone focus entirely on things that sound good, such as patriotism, they lack substance, but open themselves to less attack from opponents and critics. Should they go the other way in their speeches, and focus primarily on substance, their message is clouded and there aren’t enough details to sell it.

There are positives and negatives to each approach, and it is clear that a general mix has been discovered and utilized by the presidential office if, on average, most speeches fall under a 25% action to 75% justification metric. However, it is another metric to be utilized for comparison, because there is still a window of comparison between just how that space is used. In a speech like Clinton’s, there can be a heavy amount of focus into one category, such as the budget in Clinton’s case. Or a diaspora of plans and policies could be introduced like Obama’s speech or Bush’s speech. Or, in a case similar to Trump’s speech, the formula can be changed slightly. There can be less of a focus on actions in order to make a harder sell and include more details on individuals helped or hurt by policies.

Looking ahead there are a number of improvements that can be made for this model. First, a better method of classification needs to be developed and utilized in order for the classifications to be better trusted. Ideally, they would be done by a wider range of people from different interest groups and compared and agreed upon. The second improvement would be to include additional tuning to the models utilized so that it can be sure that the best model can be attained. Also, further cross evaluations would be important for future train and test sets so that the whole dataset can be used for training and a better idea of the model can be gleaned from something rather than chance.

Each of these speeches looked at can be their own case study into how the presidency behaves, and comparing these speeches with the whole canon of speeches at large can serve to create a clearer picture of how these decisions have changed through the years, which of these decisions were effective, and which decisions could be harmful. As such, next steps for this would be to first utilize all state of the union addresses made by each of these four presidents in order to see a clearer model, and then to include all state of the union addresses in general to compare the timeline of the past century of presidential addresses. This can then be iterated to apply towards other speeches that candidates make to hold them accountable for their words during elections. When it comes to democracy, accountability does seem to always rank as one of the highest features, and developing trust with the words that one says is one way to develop that maxim in a positive way.