Characterizing United States Presidential Candidates' Speech Patterns

Kyana Van Houten and Brendan Corcoran
Department of Computer Science, Stanford University
(Dated: June 6, 2016)

We seek to quantitatively characterize the speech patterns of leading candidates of the 2016 United States Presidential Election. We compiled full-text transcripts of stump speeches, debates, and interviews from the 2016 election cycle. Then, we determined appropriate features from the raw text that meaningfully reflect both what a candidate speaks about and how a candidate speaks. From these features, we created a model for the speech style of each candidate. Comparing these models gives insight into the differences between the candidates. It also allows us, given a new speech transcript, to make a prediction of the most likely speaker among the candidates.

INTRODUCTION

Rhetoric is a crucial factor that citizens use to decide which candidate they vote for. It affects how citizens perceive candidates on both the conscious and subconscious level, therefore it is an important topic to research and characterize. This is especially the case in the 2016 United States Presidential Election, as the candidates are perceived to have particularly different styles of speaking: from Trump's brash style to Sanders's exasperated tone. Rather than relying on stereotypes and preconceptions, we hope to rigorously define these differences by objectively characterizing the candidates' speech patterns. Furthermore, we can use these quantitative characterizations to actually predict, given a new speech transcript, which of the candidates is the most likely speaker.

Literature Review

For the reasons just described, politicians' rhetoric has been an active topic of research for several decades. One study in Sweden was able to classify politicians by gender, political affiliation, and gender [1]. Using Support Vector Machines (SVMs) and a bag-of-words vector representation, the algorithm was able to achieve an accuracy rate of 78.9% for age, 89.4% for political affiliation, and 81.2% for gender. This suggested that speech patterns varied significantly between different groups.

Another group was interested in studying the speech patterns of Japanese Prime Ministers [2]. Since SVMs do not work as well in Japanese, the group used a different method, random forest classifiers. The method successfully identified speaker-specific expressions and allowed for objective investigation of political styles.

A third group at Northwestern University examined speeches given in United States Senate [3]. Using SVMs, the algorithm achieve prediction accuracy of 92% and determine that cultural rather than economic vocabulary was more effective at differentiating liberals and conservatives.

Meanwhile, a group from the American Enterprise Institute utilized a bag-of-words technique to examine the differences between the most frequently used words by conservative and liberal 2016 Presidential candidates during debates [4]. They were able to use this technique to measure how liberal or conservative a candidate and see how this measure varied among the 2015-16 debates.

DATASET AND FEATURES

Gathering Raw Data

First, we needed to gather text data of candidates' speeches. From a number of online sources including: various news publications, the candidates' respective campaign websites, and Project Vote Smart (a nonpartisan database of information on candidates for public office) [5] [6] [7], we collected 30-40 speeches for each candidate, mostly stump speeches but also long-form responses from debates and interviews.

Feature Selection

Essentially, we want to reduce a whole speech into a feature vector. One way to do this is with a bag of words; that is, represent the speech with a vector the length of the dictionary whose i^{th} element is 1 if the ith word in the dictionary appears in the speech and 0 otherwise. This is a perfectly valid feature vector that would likely make good predictions about which candidate gave a given speech. However, this representation does not reveal anything meaningful about *how* it made the prediction. That is, we would not be able to examine the parameters and say something significant about the candidates. The best we could do is to say that, for example, Clinton is more likely to say this particular word than Trump, which in general is not very interesting.

Instead, we want our features to be significant on their own; for example, how much a candidate talks about the economy or the average word length in a speech. If we can reduce a speech to a small number of meaningful features, and we still have a lot of predictive power, this tells us much more than the bag of words approach.

What a candidate speaks about	How a candidate speaks
economy	ratio of first name-last name
health care	ratio of female-male pronouns
foreign threats	ratio of first person singular-plural
trade	mean word length
manufacturing	mean sentence length
women's health	mean number of conjunctions per sentence
immigration	number of unique words
religion	
Constitution	

Table I. Feature set

To that end, we divide our features into what a candidate talks about and how a candidate talks. In the former category, we chose 9 policy topics that we believe represent important issues as well as issues that differentiate the candidates. For each topic, we generated a list of buzzwords relating to that topic. For example, the list of words corresponding to economy includes "economy economics middle-class Wall Street banks mortgage income financial." Then, we go through each word in a speech and if that word is a buzzword of a policy, that policy is incremented. In the end, for each topic, we find the frequency (per 100 words) with which a candidate mentions a buzzword from that topic's list.

For the other category, we seek to characterize how the candidate speaks, regardless of what he or she is speaking about. This includes features such as mean sentence length and frequency of using conjunctions. The complete set of features is listed in Table I.

Data Representation

We treat a speech and its speaker as a single example. Then, the $x^{(i)} \in \mathbb{R}^k$ vector contains the real values for each of the k features in Table I for the ith training example (i^{th} speech) and $y^{(i)} \in \{1, 2, 3, 4, 5\}$ represents which of the 5 candidates gave the speech. For example, $x_4^{(10)}$ is the frequency of trade-related terms in the 10th speech. Essentially, we take a full text transcript and boil it down to a vector in $\mathbb{R}^k$.

METHODS

Algorithm Selection

There are a few options when choosing an algorithm for a multinomial classification problem. One is logistic regression (or softmax regression). This method is robust in that it does not make very strong assumptions about the data. However, since it is a discriminative algorithm, its parameters do not tell us what a Clinton or Trump speech looks like on its own, but only a relative probability between the two.

Another option is Gaussian discriminant analysis (GDA). This makes stronger assumptions about the data, namely that the data is indeed Gaussian. Our feature set contains features with distributions that are distinctly Gaussian (mean word length, number of unique words) and some features whose distributions are more Poisson (frequency of policy buzzwords). However, we feel comfortable approximating every feature as a Gaussian and trusting that GDA is robust enough to make good predictions (this proved to be true in practice).

Another advantage of GDA is that it tends to require fewer training examples to learn well. One of our major limitations is sheer number of speeches a candidate gives as well as the availability of transcripts. As a result, we managed to collect 30-40 speeches per candidate which is not a very large training set. Despite this, GDA should be able to generate good results.

Generating Parameters

Following the GDA strategy, we fit a k-dimensional Gaussian to each of the candidates' training data. To do this, we must determine the mean vector (μ) for each candidate as well as one covariance matrix (Σ) that will

be used for all candidates. The parameters are determined by the following intuitive equations, where m is the size of the training set:

$$\mu_j = \frac{\sum_{i=1}^{m} 1(y^{(i)} = j)x^{(i)}}{\sum_{i=1}^{m} 1(y^{(i)} = j)}$$

$$\Sigma = \sum_{i=1}^{m} (x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^T$$

Making a Prediction

Given a new speech, we want to predict which candidate is most likely to have given that speech. That is,

$$y_{pred} = \arg\max_y p(y|x)$$

However, we only know $p(x|y)$; namely the Gaussian:

$$p(x|y) \sim \exp(-\frac{1}{2}(x - \mu_y)^T \Sigma^{-1}(x - \mu_y))$$

We can relate the two through Bayes rule to get:

$$y_{pred} = \arg\max_y \exp(-\frac{1}{2}(x - \mu_y)^T \Sigma^{-1}(x - \mu_y))p(y)$$

This is how the algorithm makes a prediction. Intuitively, we find which candidate's Gaussian the new point is most likely to lie on.

RESULTS AND ANALYSIS

There are several results we wanted to obtain from this study: the ability to predict which candidate was most likely to have given a speech, the ability to compare meaningful statistics about the candidates' speech, and the ability to classify historical speeches to current candidates.

Predictive Power

To make a prediction, we pre-process the new speech example as we do the training data to create a feature vector. Using this feature vector, we determine which candidate's Gaussian the new example is most likely to lie on then compare the result to the identity of the known orator. Figures 1 and 2 show examples of the models constructed for the candidates in two different feature-spaces. In determining the accuracy with which our method can predict which candidate was the orator of a speech, we randomly shuffled and split our collected

speeches into sets of 80% training and 20% test sets. We shuffled and split our speeches 10 times to train and test our models. Averaging across these trials, we achieved a training error of 29% and a test error of 31% both of which are well below the expected error from random guessing (80%).
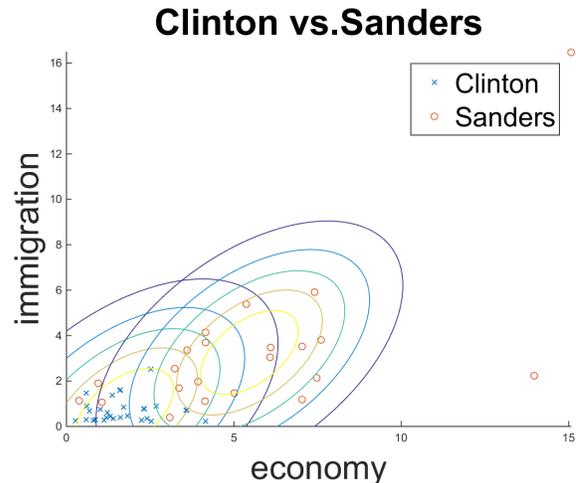

Figure 1. Comparison of the distributions of frequency with which Hillary Clinton and Bernie Sanders speak about immigration and the economy. We see that Sanders speaks more about the economy and immigration than does Clinton.
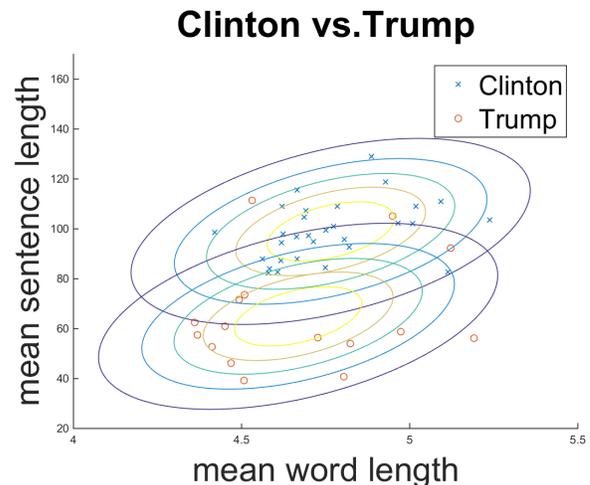

Figure 2. Comparison of the distributions of mean word length and mean sentence length in the speeches from Hillary Clinton and Donald Trump We see that Trump generally uses slightly shorter words than Clinton on average while Trump's sentences are significantly shorter than Clinton's on average.

Comparing Candidates

We also measured several interesting results about each candidate's speech patterns. For example, Bernie

Sanders spoke the most about the economy. Ted Cruz spoke the most about religion, the Constitution, and foreign threats. Donald Trump spoke the most about immigration and had the shortest mean sentence length. Trump tends to use first person singular tense significantly more than Cruz. The two Democrat candidates tend to speak more about manufacturing than the Republican candidates.

Figure 3 allows us to compare all of the candidates in terms of all of the features we chose. We can also generate figures similar to Figures 1 and 2 for any pair of candidates with respect to any features of interest for a more in-depth comparison of the candidates.
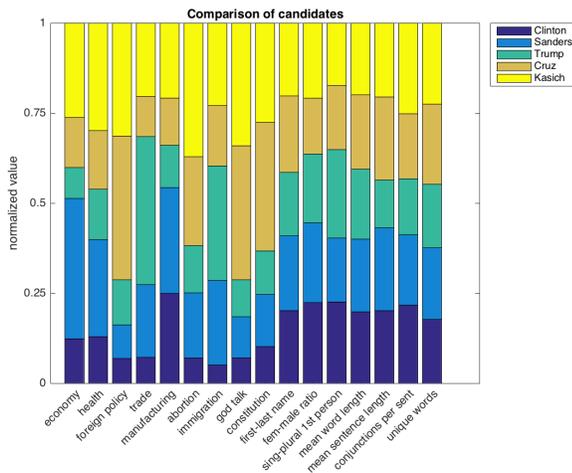

Figure 3. Relative mean values for each candidate for each feature

Historical Speeches

Once we have tested our algorithm on speeches we know to have come from the candidates, we thought it would be interesting to feed the algorithm famous speeches from historical figures to determine which of the present-day candidates would be most likely to give the speech.

From this testing, we found that some of the speeches our model attributed to Ted Cruz were Barry Goldwater's acceptance speech at the 1964 Republican National Convention as well as President Reagan's speech on the "Evil Empire."

Among those predicted to be Hillary Clinton speeches were John F. Kennedy's moon speech at Rice and Obama's victory speech on the night of the 2012 general election.

However, interestingly enough, Bill Clinton's speech at the 2012 Democratic National Convention was predicted to be from Bernie Sanders.

CONCLUSION

Our project demonstrated the effectiveness of an alternative method to the commonly used bag-of-words for characterizing speech patterns by representing speeches only as a vector of feature values instead of much larger word-vectors. Our method also provides us with much richer characterization as opposed to just prediction capabilities.

In the future, we can add additional features that will allow us to analyze speeches in more depth. We were limited to features that were fairly simple to extract, but with more background in natural language processing we could expand our features to include measures of the sophistication of language used by candidates (i.e. Candidate 1 speaks at a 7th grade comprehension level whereas Candidate 2 speaks at a university comprehension level) which tells us how accessible a candidate may be to different demographics. We could also add features that extract a candidate's position on different issues rather than just how often a candidate speaks about an issue.

This method has potential to substantially contribute to the data-driven aspect of politics and allow voters to compare candidates in a quantitative manner.

[1] Mats Dahllöf. Automatic prediction of gender, political affiliation, and age in swedish politicians from the wording of their speeches—a comparative study of classifiability. *Literary and Linguistic Computing*, 2012.

[2] Takafumi Suzuki. Extracting speaker-specific functional expressions from political speeches using random forests in order to investigate speakers' political styles. *Journal of the American Society for Information Science and Technology*, 2009.

[3] Bei Yu Daniel Diermeier, Jean-François Godbout and Stefan Kaufmann. Language and ideology in congress, 2011.

[4] Weifeng Zhong. The candidates in their own words: A textual analysis of 2016 presidential primary debates. *AEI Economic Perspectives*.

[5] *The New York Times*.

[6] *Project Vote Smart*.

[7] *Hillary Clinton Campaign Website*.