

Political Ideology Detection

Dong Wook (Daniel) Chung '17

TRA301 – Introduction to Machine Translation

Professor Bangalore

May 6th 2016

Motivation

1. Media Bias

- Word Choice
 - Abortion: “Pro-life (anti-abortion)” vs “Pro-choice (pro-abortion)”
 - Tax: “estate tax (liberals)” vs “death tax (conservatives)”
- Some media are accused of favoring one side:

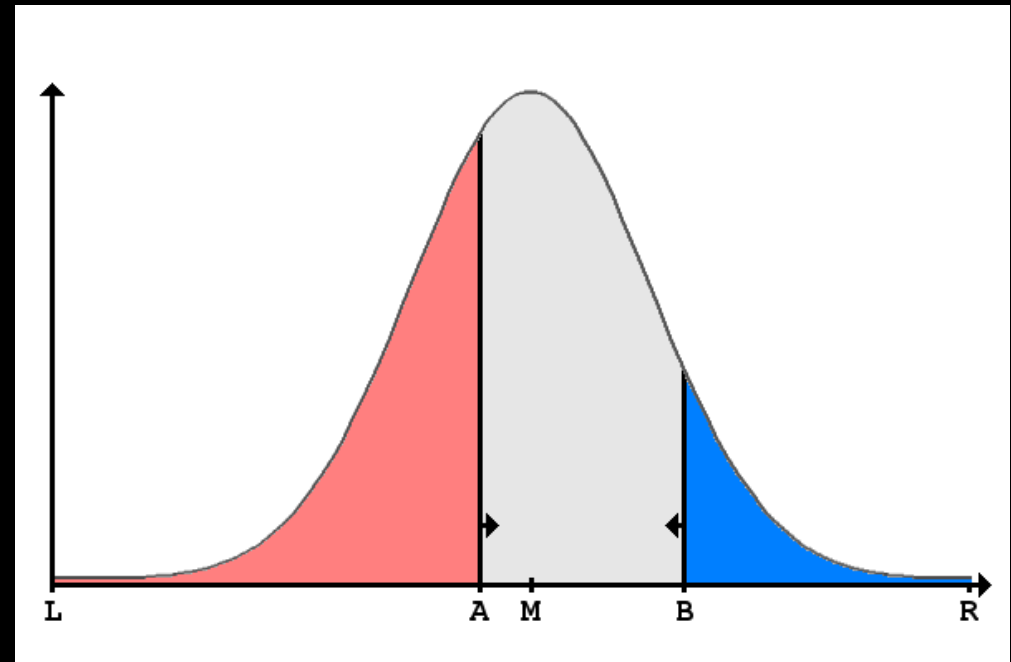


- Problem: it is difficult to understand politicians' ideologies correctly through media

Motivation

2. Median Voter Theorem (MVT) – is it true?

- Theorem: “A majority rule system will select the outcome most preferred by the median voter”
- Assumptions
 - A one-dimensional political spectrum
 - Single-peaked



Goal

1. Identify politicians' ideologies from their speeches
2. Test Median Voter Theorem
 - Hypothesis: in presidential elections, candidates tend to be more “neutral” in general election than in primary election
 - In general election, voters' political ideologies are more widely spread out

Method



1. Political ideology detection:

- Target: 2008, 2012 presidential election candidates
- Each speech is classified as “democratic” or “republican” using text classifier
- For each candidate, compute the ratio

The number of “democratic” speeches

The total number of speeches

2. MVT

- For Obama, McCain and Romney, compare the distribution of their speeches between primary election and general election

Data

- 2005 Congress speeches –for training text-classifier
 - 4046 speeches from the Democratic Party
 - 4075 speeches from the Republican Party
- 2008, 2012 presidential candidates' speeches –for testing
 - 18 candidates
 - 809 speeches

Text-classification

1. Feature Extraction

- Word frequency matrix
- Converted into tf-idf matrix
- Tf-idf (term frequency – inverse document frequency): in a collection of documents, measures how important a word is to a given document

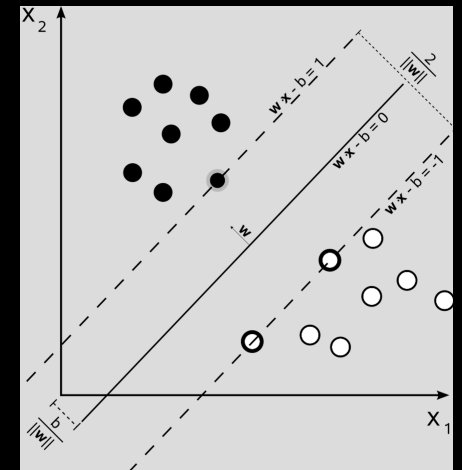
2. Training text-classifiers

- Naïve Bayes

$$\hat{y} = \arg \max_y P(y) \prod_{i=1}^n P(x_i | y) \quad \text{where } y = \{\text{Democrat, Republican}\}$$

- Assumption: each word is independent of other words in the same document

- Linear SVM



Text-classification

- 10-fold cross-validation
- Tested unigram, bigram, trigram, quadgram and their combinations
- Best F1 score after some improvements:
- Linear SVM using unigram, bigram and trigram
 - The Democratic Party: 0.7721
 - The Republican Party: 0.8028
 - Average: 0.7874

The Democratic Party candidates result

(0% = completely Republican, 100% = completely Democratic)

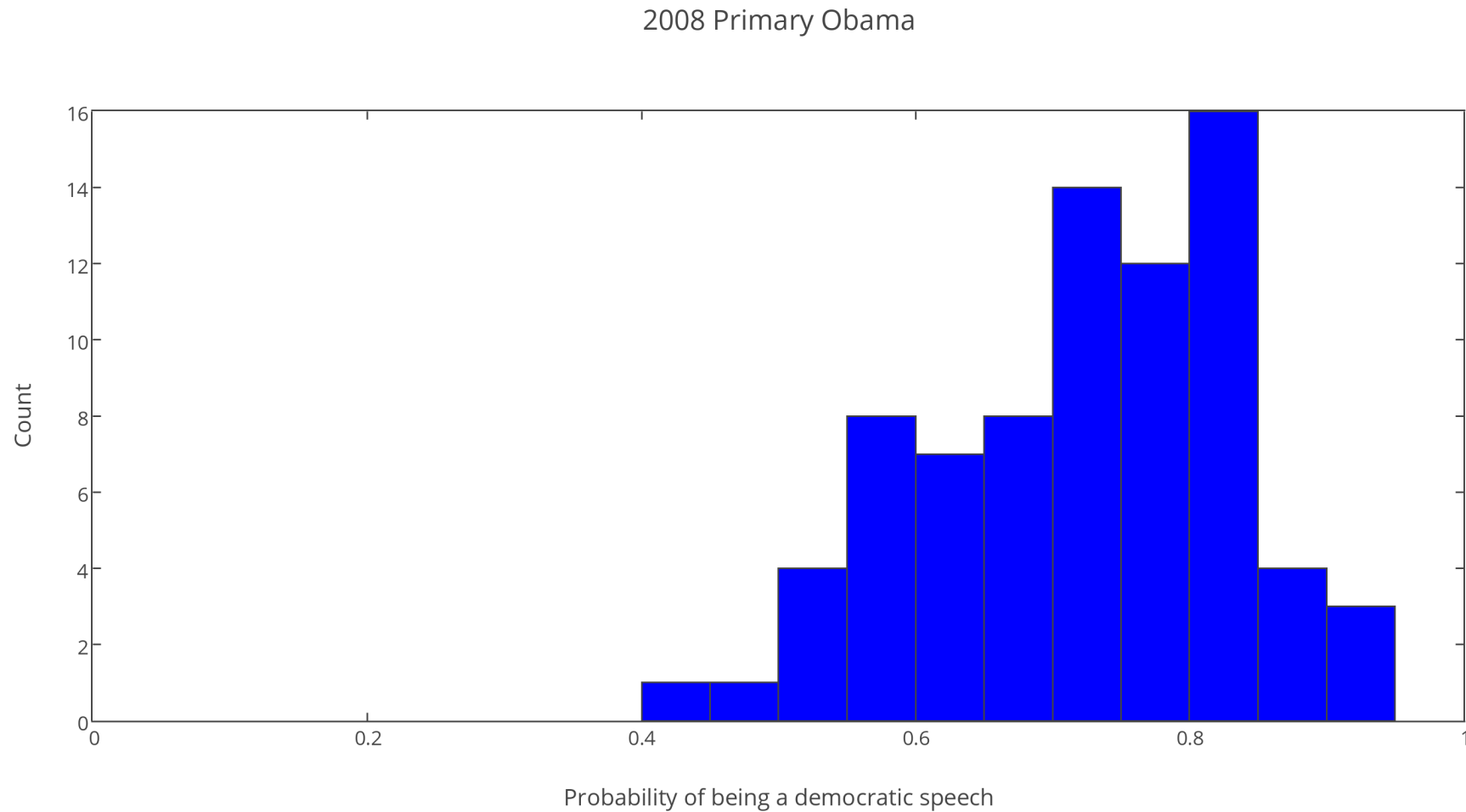
Name	Republican	Democratic	Total	Ratio	OnTheIssues
John Edwards	25.7	1.3	27	95.2%	Populist-leaning liberal
Hillary Clinton	92.1	13.9	106	86.9%	Hard-core liberal
Barack Obama	212.5	59.5	272	78.1%	Hard-core liberal
Bill Richardson	22.5	8.5	31	72.6%	Populist-leaning liberal

The Republican Party Candidates Result

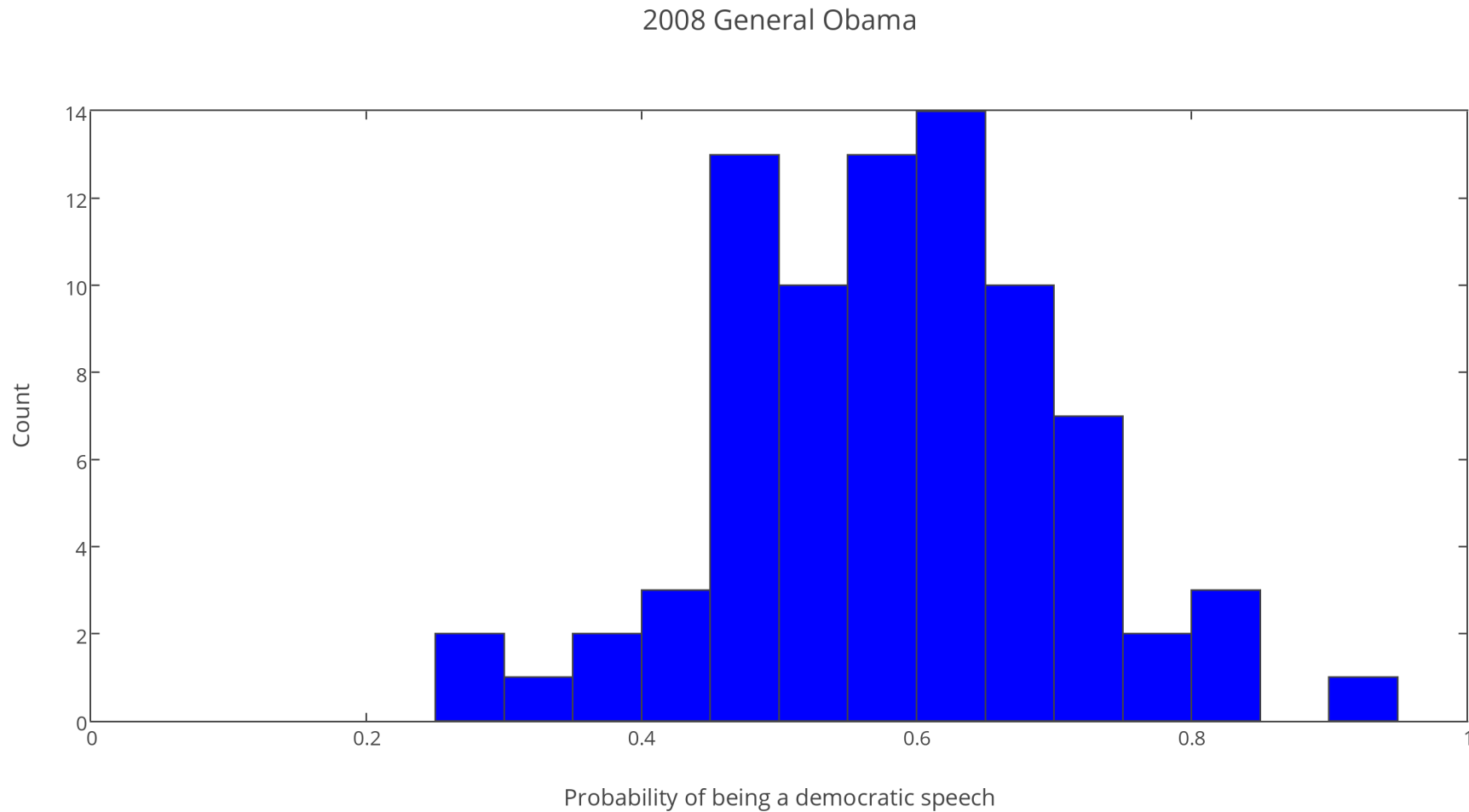
(0% = completely Republican, 100% = completely Democratic)

Name	Republican	Democratic	Total	Ratio	OnTheIssues
Newt Gingrich	1.8	13.2	15	12.0%	Hard-core conservative
Rick Perry	1.4	8.6	10	14.0%	Hard-core conservative
Rick Santorum	3	13	16	18.8%	Hard-core conservative
Rudy Giuliani	9.7	27.3	37	26.2%	Libertarian conservative
Mike Huckabee	5.2	9.8	15	34.7%	Populist-leaning conservative
Mitt Romney	23.4	37.6	61	38.4%	Populist-leaning conservative
Fred Thompson	6.7	9.3	16	41.9%	Moderate conservative
John McCain	84.3	89.7	174	48.4%	Libertarian conservative

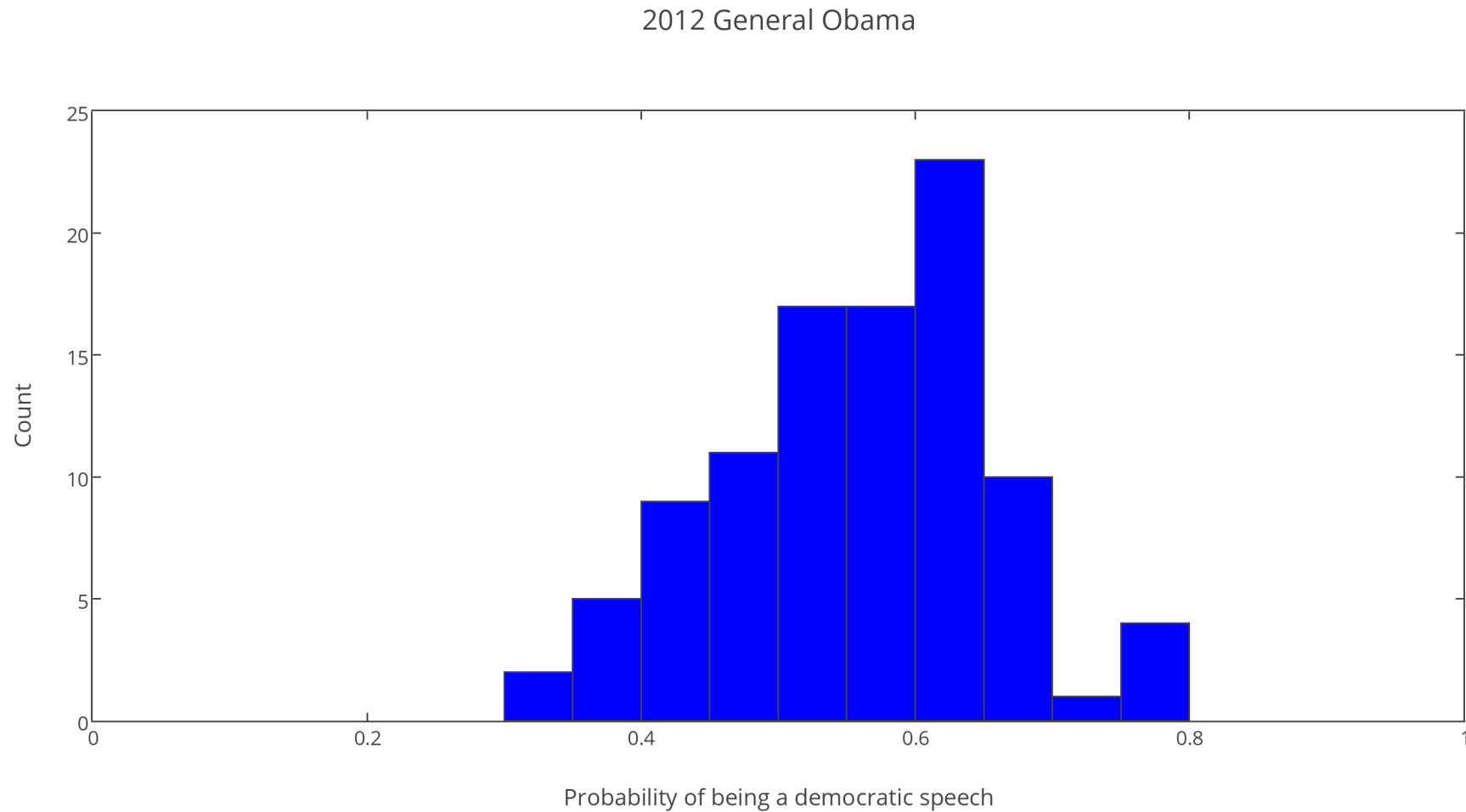
Obama 2008 Primary (72.0%)



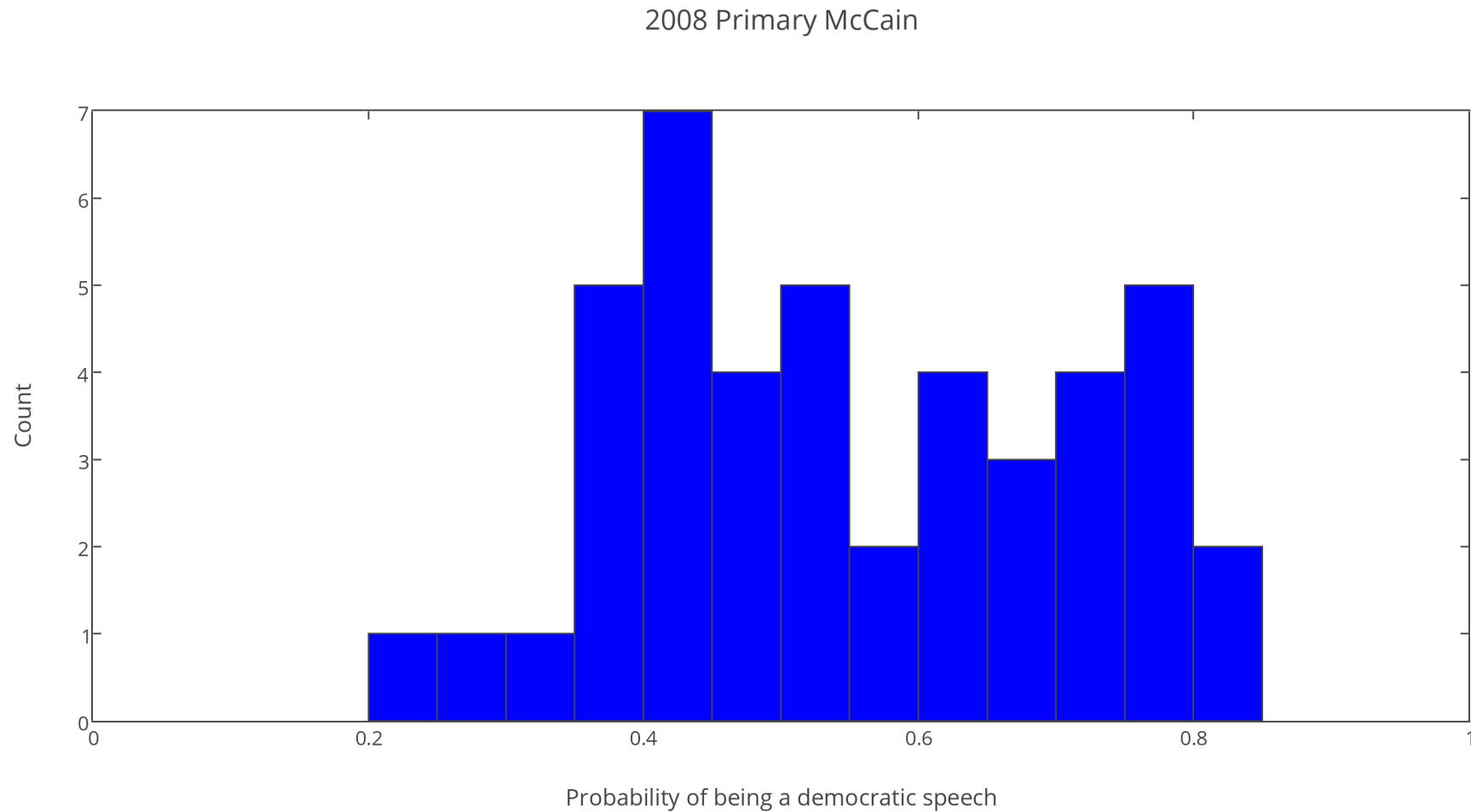
Obama 2008 General (58.4%)



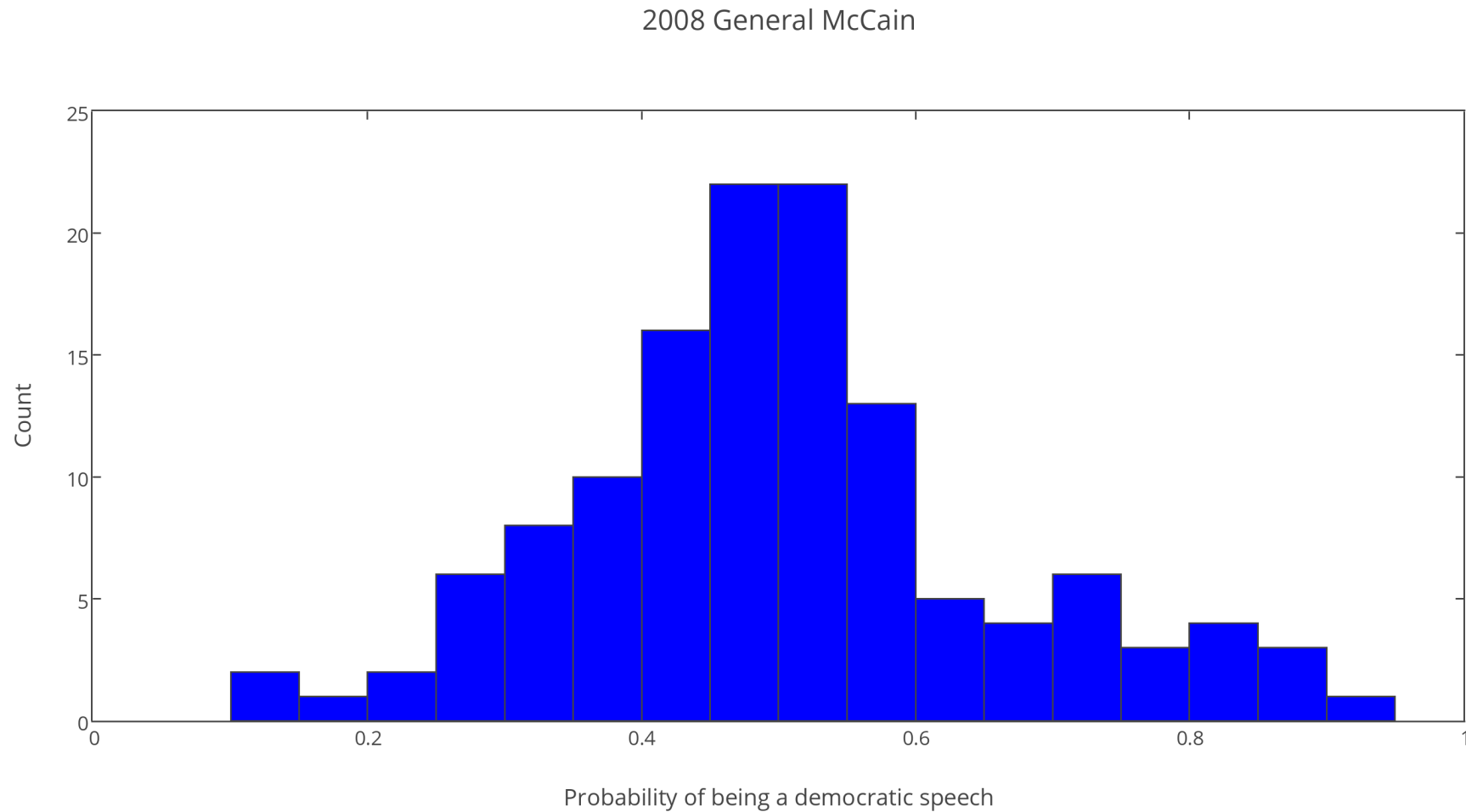
Obama 2012 General (56.1%)



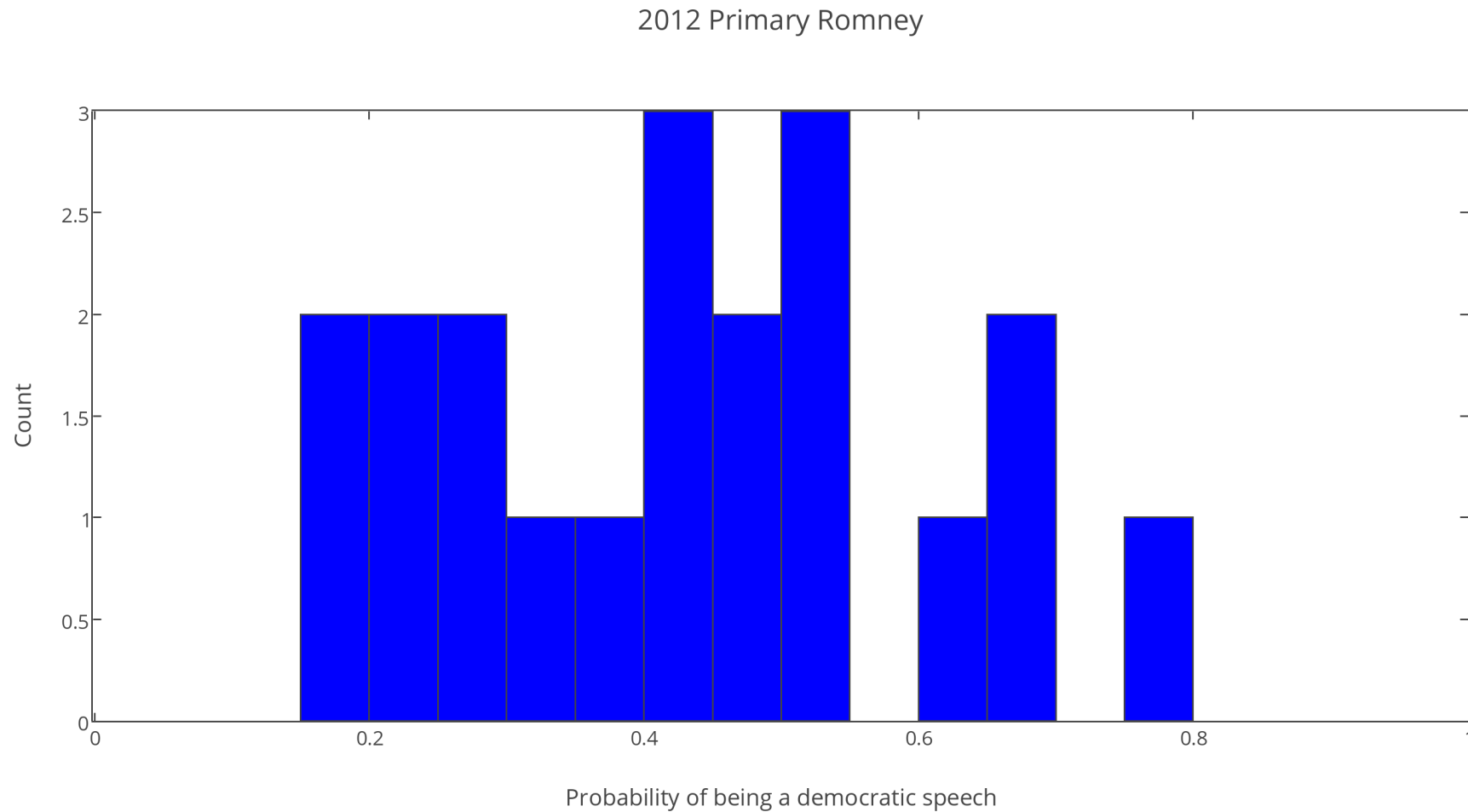
McCain 2008 Primary (55.1%)



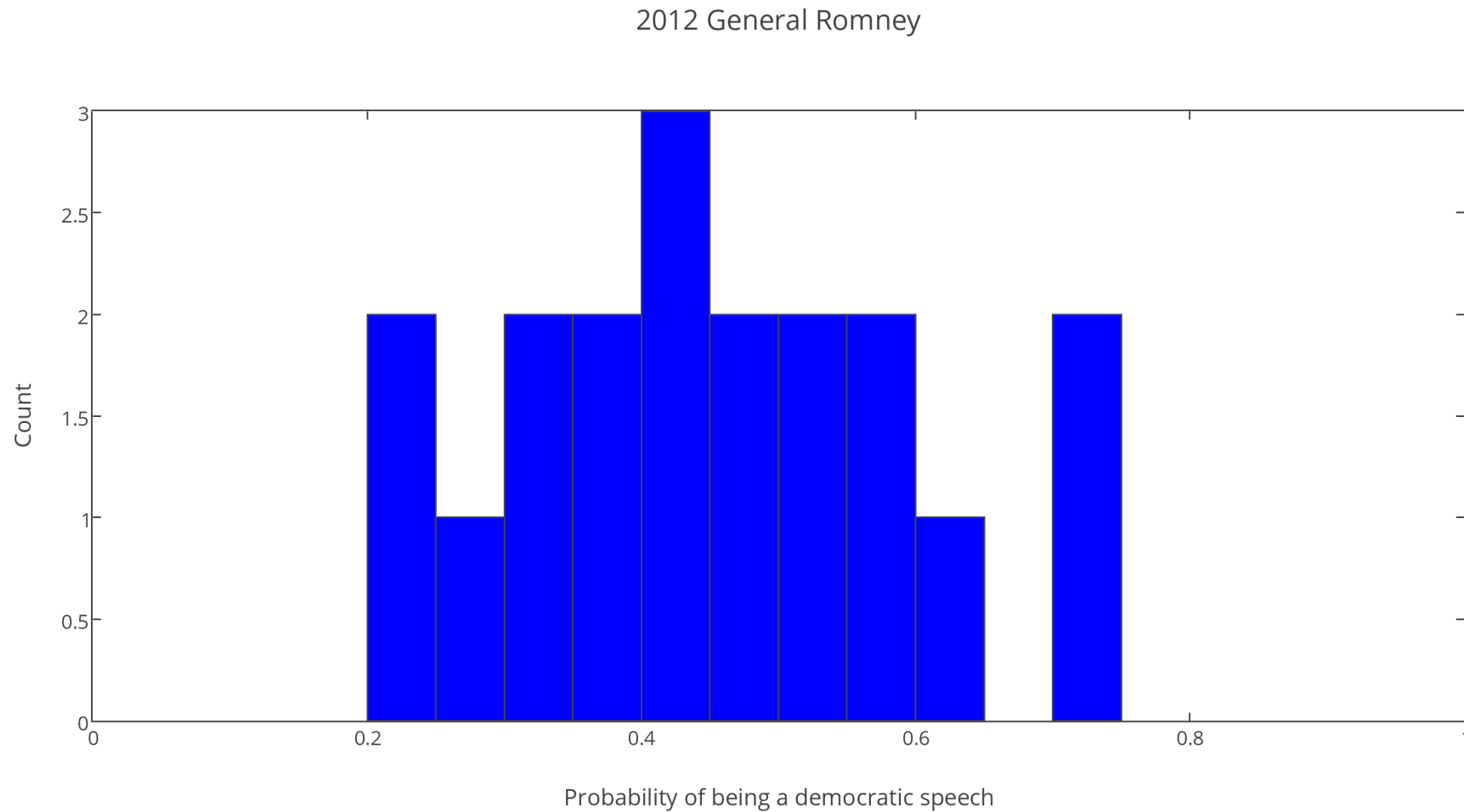
McCain 2008 General (50.0%)



Romney 2012 Primary (43.6%)



Romney 2012 General (44.9%)



Future Work

- More political speeches for text-classification training -> improved accuracy
- Test MVT with more candidates from past elections
- Understand what makes a speech classified as democratic or republican
- Automatic Political Ideology Generator
 - Given a politician's name and a keyword (e.g. Obama, Wall Street), summarize the politician's view on the given topic
 - Focus on the words and phrases near the keyword

Conclusion

- Identified a politician's ideology using speeches and text-classifiers
- Tested MVT using the speeches from Obama, McCain and Romney



Donald Trump's Political Ideology

Name	Republican	Democratic	Total	Ratio	On the Issues
Donald Trump	12.7	23.3	36	35.3%	Moderate Conservative

From The Washington Post:

“Donald Trump is a textbook example of an ideological moderate”