



# Sentiment Analysis of Political Statements Using Cross-Field Training Corpora

Krissy Gianforte  
MIDS w266, Fall 2018

# Can Official Online statements be Used to Predict Congressmen's Votes on an Issue?


.gov websites Healthcare issue statements

About Contact Services Media Center Legislation Issues Resources

UNITED STATES CONGRESSWOMAN  
**ALMA ADAMS**  
for the Twelfth District of NORTH CAROLINA

Home » Issues

## Health Care



Health care is important to our district and to Congresswoman Adams' work in Congress. She has been a staunch and ardent supporter of the Affordable Care Act, which has brought lifesaving insurance coverage to millions, including hundreds of thousands in North Carolina's 12th District.

Congresswoman Adams is a proud cosponsor of bills that protect women's health, medical research and the elimination of taxes on health care benefits for Middle Class families. She has also written to North Carolina Governor Pat McCrory urging that he expand Medicaid in the state for more than 319,000 North Carolinians.

Votes on HR1628, to repeal the ACA

### American Health Care Act of 2017



**Acronym** AHCA

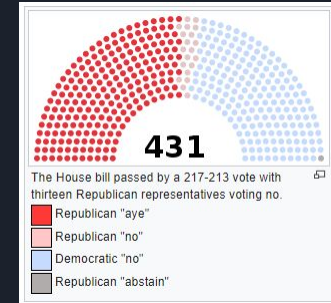
**Colloquial name(s)** Trumpcare, Ryancare, Republicare, Obamacare Lite

**Introduced in** 115th United States Congress

**Introduced on** March 20, 2017

**Effects and codifications**

**Act(s) affected** Patient Protection and Affordable Care Act, Medicare Access and CHIP Reauthorization Act of 2015, Social Security Act, Balanced Budget and Emergency Deficit Control Act of 1985, Public Health Service Act





# Previous Work: Stance Analysis of Tweets

*Kristen Johnson & Dan Goldwasser* explored stance and consistency in position over time

Collected 99,161 tweets for 32 politicians, covering 16 issues

## KEY IDEAS:

- LOTS of data!
- Filtered tweets by topic keywords
- Used **OpinionFinder2.0** to label sentiment
  - Includes a **Subjectivity Analysis** filter to extract “internal mental or emotional states, including speculations, beliefs, emotions, evaluations, goals, and judgements”



# Previous Work: Stance Analysis of Tweets

*Saif Mohammad, Parinaz Sobhani, & Svetlana Kiritchenko* explored stance using a SVM

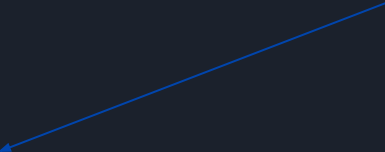
$\text{STANCE} \sim \text{target} + \text{encodings} + \text{sentiment} + \text{n-grams} + \text{POS}$



# Previous Work: Stance Analysis of Tweets

*Saif Mohammad, Parinaz Sobhani, & Svetlana Kiritchenko* explored stance using a SVM

STANCE ~ target + encodings + sentiment + n-grams + POS



Filtering

# Previous Work: Stance Analysis of Tweets

Saif Mohammad, Parinaz Sobhani, & Svetlana Kiritchenko explored stance using a SVM

STANCE ~ target + encodings + sentiment + n-grams + POS

Filtering

*Emoticons, hashtags,  
uppercase, elongated words,  
extreme punctuation*  
Not relevant for composed,  
professional statements

# Previous Work: Stance Analysis of Tweets

Saif Mohammad, Parinaz Sobhani, & Svetlana Kiritchenko explored stance using a SVM

STANCE ~ target + encodings + sentiment + n-grams + POS

Filtering

*Emoticons, hashtags,  
uppercase, elongated words,  
extreme punctuation*  
Not relevant for composed,  
professional statements

Limited dataset  
(410 votes;  
135 statements)

# Previous Work: Stance Analysis of Tweets

Saif Mohammad, Parinaz Sobhani, & Svetlana Kiritchenko explored stance using a SVM

STANCE ~ target + encodings + sentiment + n-grams + POS

Filtering

*Emoticons, hashtags,  
uppercase, elongated words,  
extreme punctuation*  
Not relevant for composed,  
professional statements

Used known  
lexicons

Limited dataset  
(410 votes;  
135 statements)



# Previous Work: Stance Analysis of Tweets

Saif Mohammad, Parinaz Sobhani, & Svetlana Kiritchenko explored stance using a SVM

STANCE ~ target + encodings + sentiment + n-grams + POS

Filtering

*Emoticons, hashtags,  
uppercase, elongated words,  
extreme punctuation*  
Not relevant for composed,  
professional statements

Used known  
lexicons

**FOCUS HERE**

Limited dataset  
(410 votes;  
135 statements)



# Can Cross-Field Lexicons Provide Adequate Training Data?

**VADER**: designed for social media text; accounts for capitalization and punctuation as intensifiers and can interpret common emoji

Evaluates a full text as positive or negative -- returns a + or - intensity measure



# Can Cross-Field Lexicons Provide Adequate Training Data?

**VADER:** designed for social media text; accounts for capitalization and punctuation as intensifiers and can interpret common emoji

Evaluates a full text as positive or negative -- returns a + or - intensity measure

**Opinion Lexicon (Liu & Hu):** based on customer product reviews

Matches words to a list of 6800 positive or negative words -- returns a word count difference



# Can Cross-Field Lexicons Provide Adequate Training Data?

**VADER:** designed for social media text; accounts for capitalization and punctuation as intensifiers and can interpret common emoji

Evaluates a full text as positive or negative -- returns a + or - intensity measure

**Opinion Lexicon (Liu & Hu):** based on customer product reviews

Matches words to a list of 6800 positive or negative words -- returns a word count difference

**SentiWordNet:** provides a sentiment score for each word in the SentiWordNet set

Sums the sentiment scores of each word in a passage -- returns a + or - total score



# Determining Sentiment

## (1) Preprocessing

Break sentences on newline and punctuation

Filter by keywords -- ["obamacare", "obama care", "affordable care", "aca"]

(considered "insurance", "coverage", "health plan", and "repeal", but the net got too wide)



# Determining Sentiment

## (1) Preprocessing

Break sentences on newline and punctuation

Filter by keywords -- ["obamacare", "obama care", "affordable care", "aca"]

(considered "insurance", "coverage", "health plan", and "repeal", but the net got too wide)

## (2) Sentiment scoring

Score statements based on the 3 lexicons



# Determining Sentiment

## (1) Preprocessing

Break sentences on newline and punctuation

Filter by keywords -- ["obamacare", "obama care", "affordable care", "aca"]

(considered "insurance", "coverage", "health plan", and "repeal", but the net got too wide)

## (2) Sentiment scoring

Score statements based on the 3 lexicons

## (3) Subjectivity filtering

Filter statements for subjective sentences only

Naive Bayes model using NLTK Subjectivity set as training data



# Determining Sentiment

## (1) Preprocessing

Break sentences on newline and punctuation

Filter by keywords -- ["obamacare", "obama care", "affordable care", "aca"]

(considered "insurance", "coverage", "health plan", and "repeal", but the net got too wide)

## (2) Sentiment scoring

Score statements based on the 3 lexicons

## (3) Subjectivity filtering

Filter statements for subjective sentences only

Naive Bayes model using NLTK Subjectivity set as training data

## (4) Sentiment scoring (again)



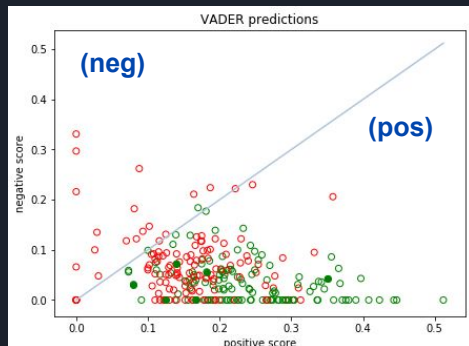
# Results

	Baseline Majority Class	VADER	Opinion Lexicon (Liu-Hu)	SentiWordNet
Voting Data	51.6%	---	---	---
Full Topic Statements	---	54.8%	57.6%	58.4%
Subjective Statements	---	55.3%	58.5%	61.3%

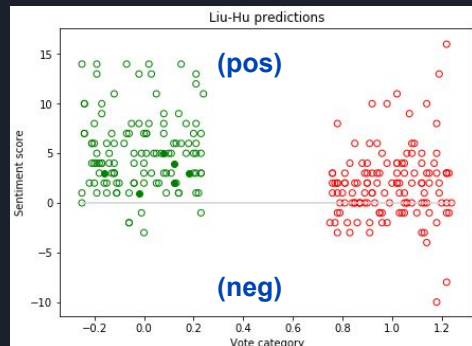
YES vote / neg sent (prediction) NO vote / pos sent (prediction)

Full  
Statements

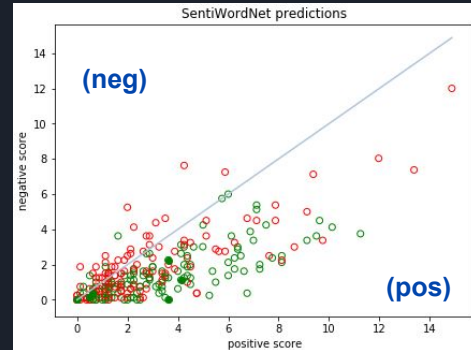
VADER



Opinion Lexicon

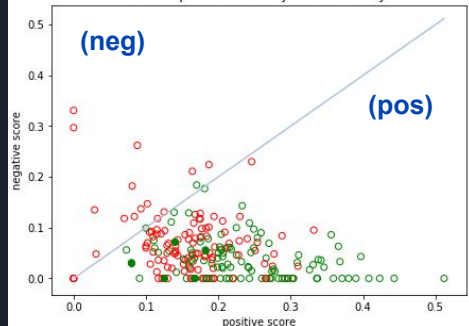


SentiWordNet

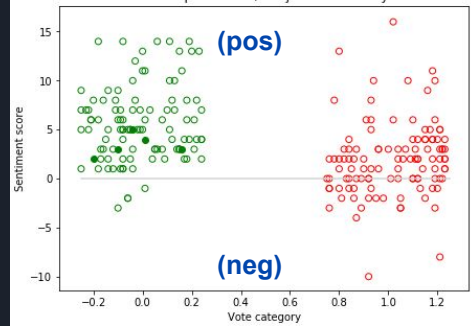


Subjective  
Statements

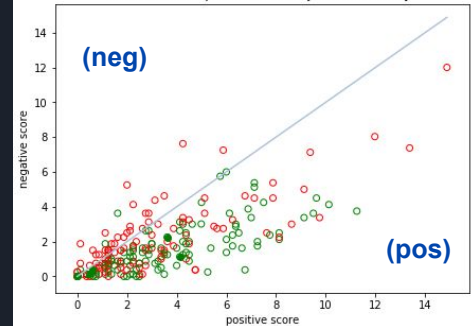
VADER predictions, Subj sentences only



Liu-Hu predictions, Subj sentences only

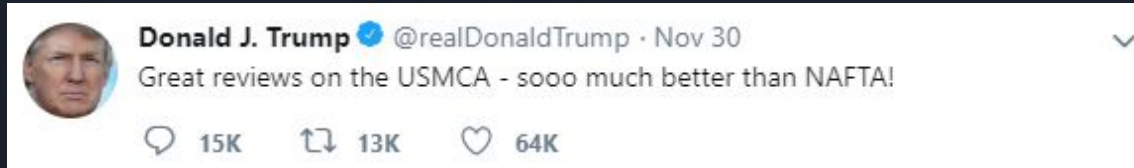


SentiWordNet predictions, Subj sentences only



# Error Patterns & Issues - POSITIVE BIAS

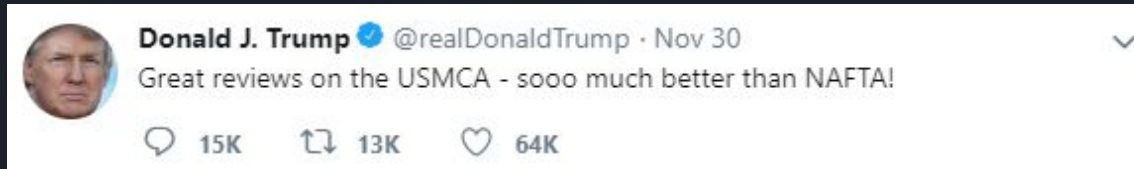
Official statements are not as expressive as twitter or product reviews



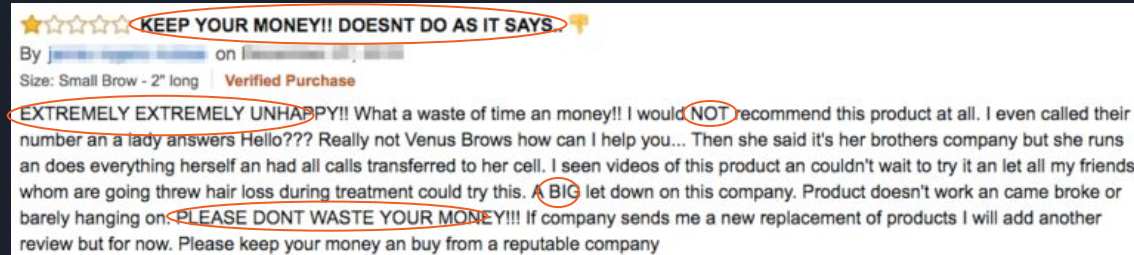
Government-centered health care models like Obamacare are fundamentally unworkable and unfair, and fail to respect the dignity of human life. Congress needs to focus its energy on practical health care solutions that lower cost and increase personal choice.

# Error Patterns & Issues - POSITIVE BIAS

Official statements are not as expressive as twitter or product reviews



Capitalization



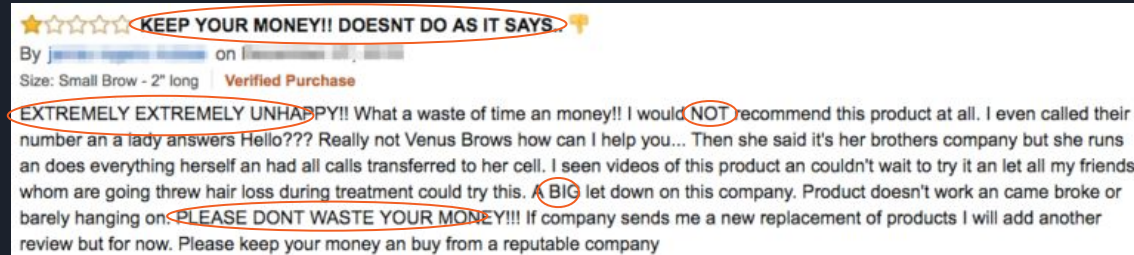
Government-centered health care models like Obamacare are fundamentally unworkable and unfair, and fail to respect the dignity of human life. Congress needs to focus its energy on practical health care solutions that lower cost and increase personal choice.

# Error Patterns & Issues - POSITIVE BIAS

Official statements are not as expressive as twitter or product reviews



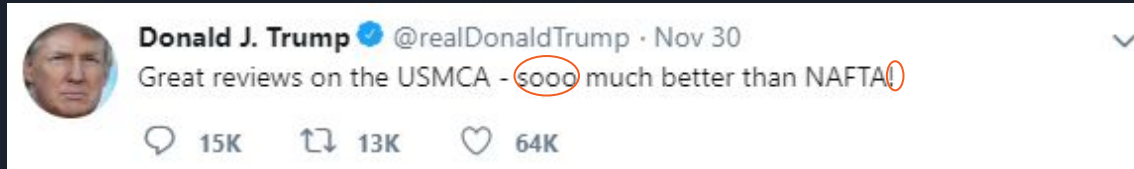
Capitalization  
Repeated letters



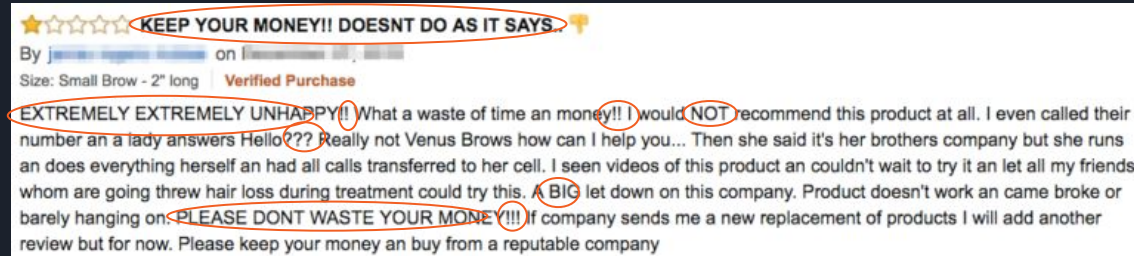
Government-centered health care models like Obamacare are fundamentally unworkable and unfair, and fail to respect the dignity of human life. Congress needs to focus its energy on practical health care solutions that lower cost and increase personal choice.

# Error Patterns & Issues - POSITIVE BIAS

Official statements are not as expressive as twitter or product reviews



Capitalization  
Repeated letters  
Punctuation

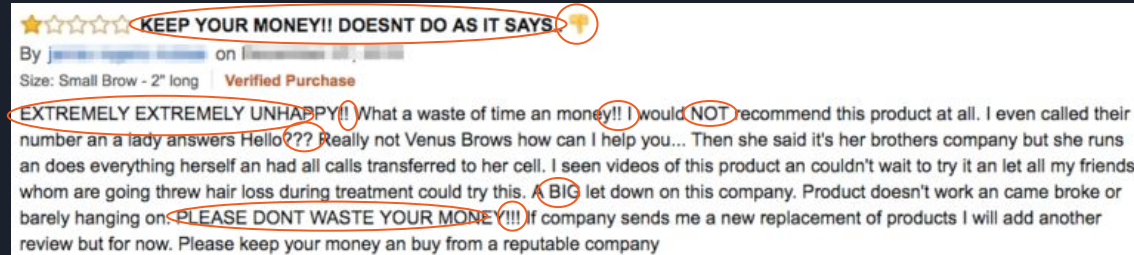


Government-centered health care models like Obamacare are fundamentally unworkable and unfair, and fail to respect the dignity of human life. Congress needs to focus its energy on practical health care solutions that lower cost and increase personal choice.



# Error Patterns & Issues - POSITIVE BIAS

Official statements are not as expressive as twitter or product reviews



Capitalization  
Repeated letters  
Punctuation  
Emoji

Government-centered health care models like Obamacare are fundamentally unworkable and unfair, and fail to respect the dignity of human life. Congress needs to focus its energy on practical health care solutions that lower cost and increase personal choice.



# Error Patterns & Issues - SUBJECTIVITY

Difficult to detect subjective statements -- everything is stated as fact

*Subjective*

*Objective...?*

*i promised the people of the 36th district of texas that i would repeal obamacare and replace it with a workable, affordable solution which lowers premiums while protecting those with preexisting conditions. while republicans alone provided the votes to pass this bill, it was still the product of finding common ground to provide the american people with relief from obamacare. unfortunately, the senate defeated this bill by a single vote, and the fight to repeal obamacare continues*

*... obamacare failed to accomplish real reform, and instead harmed health care, job creation, and the federal deficit at a time when our country could not afford such government inflicted damage. obamacare caused premiums to skyrocket, forcing millions of americans off of their current coverage and putting unelected washington bureaucrats between patients and their doctors. obamacare's equation is simple: higher costs, more debt, fewer doctors, reduced access, fewer jobs, and increased dependency on failed federal programs.*



# Final Conclusion

**Relevant**, abundant training data is critical for accurate modeling





# Final Conclusion

**Relevant**, abundant training data is critical for accurate modeling

A Naive Bayes model using these sentiment scores but *trained to predict votes based on the actual dataset*

$VOTE \sim VADER\_pos + VADER\_neg + Opinion\_sent + SentiWordNet\_pos + SentiWordNet\_neg$



# Final Conclusion

**Relevant**, abundant training data is critical for accurate modeling

A Naive Bayes model using these sentiment scores but *trained to predict votes based on the actual dataset*

$$\text{VOTE} \sim \text{VADER\_pos} + \text{VADER\_neg} + \text{Opinion\_sent} + \text{SentiWordNet\_pos} + \text{SentiWordNet\_neg}$$
 achieves 66% accuracy!



# Final Conclusion

**Relevant**, abundant training data is critical for accurate modeling

A Naive Bayes model using these sentiment scores but *trained to predict votes based on the actual dataset*

$VOTE \sim VADER\_pos + VADER\_neg + Opinion\_sent + SentiWordNet\_pos + SentiWordNet\_neg$

achieves 66% accuracy!

Politics is a strange beast...

Republican Congressman Biggs voted AGAINST the bill (against repealing the ACA)

*Obamacare has become a pariah for all who helped with its passage, and has led to unaffordable premiums and fewer options. ... Since taking office at the start of the 115th Congress, I have advocated for the immediate passage of a bill that completely repeals Obamacare and all the regulations associated with it. ... I am committed to the challenge of eliminating Obamacare and to re-establishing our nation's healthcare system to once again be the envy of the world.*

Votes are cast for many reasons, not always concretely related to ideology.