Biased Language in News Media

Tom Wang

Contents

- I. Intro
- II. News Websites Classifier
 - III. News Topic Modeling
- IV. Biased Language in the News
 - V. Future Work / Questions

About

II. News Website Classifie

Data Scientist

Interested in how bias can influence language and the news

III. News Topic Modeling • Question: Can NLP techniques detect differences in language used when comparing different news media sources?

IV. Biased Language Wiki

The Datasets

II. News Website Classifier

III. News
Topic
Modeling

IV. Biased Language Wiki

- News Articles Dataset
 - ~140k News articles scraped from 15 different news media websites
 - o 2015 2017
- Bias Lexicon
 - List of 654 bias-inducing lemmas
 - Stanford researchers derived from Wikipedia edits corpus

MediaBiasFactCheck.com

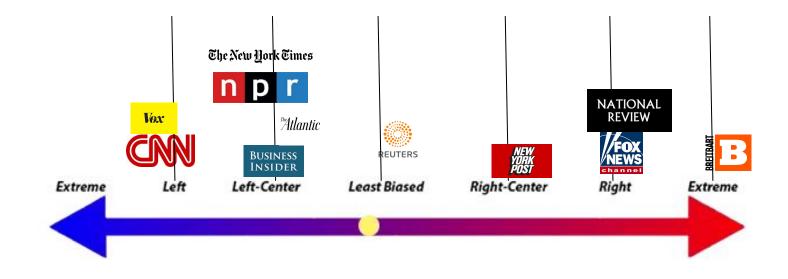
II. News Website Classifier

III. News Topic Modeling

IV. Biased Language -

IV. Future / Questions

"Media Bias/Fact Check (MBFC), founded in 2015, is an independent online media outlet. MBFC is dedicated to educating the public on media bias and deceptive news practices." [1]



II. News Website Classifier

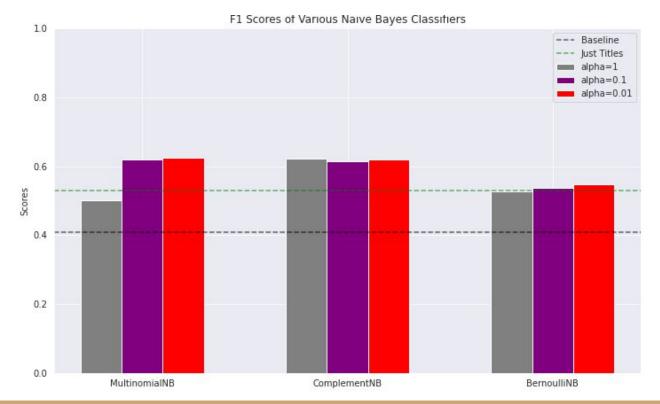
III. News Topic Modeling

IV. Biased Language -Wiki

IV. Future / Questions

Naive Bayes Classifiers - F1 Score

- TF-IDF Vectorized article text
- Multinomial/Bernoulli/Complement Naive Bayes

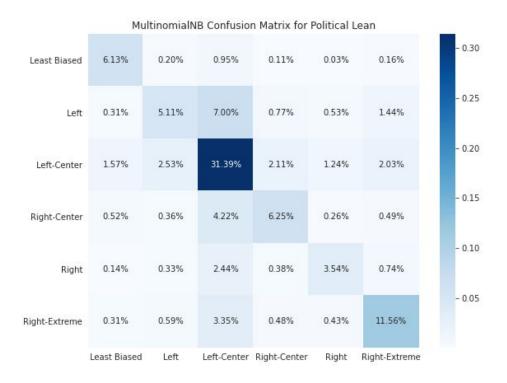


Naive Bayes Confusion Matrix

II. News Website Classifier

III. News Topic Modeling

IV. Biased Language -Wiki



Topic Modeling on News Set

Unsupervised Feature Generation - Topics / Subject Matter

Methodologies

- Gensim Latent Dirichlet Allocation (LDA)
- NMF Non-negative Matrix Factorization

Topic **Modeling**

III. News

Model	Training time (seconds)
LDA	334.6
NMF	3.6

Topic Modeling on News Set

II. News Website Classifie

Unsupervised Feature Generation - Topics / Subject Matter

Methodologies

III. News
Topic
Modeling

- 1. Gensim Latent Dirichlet Allocation (LDA)
- 2. NMF Non-negative Matrix Factorization

IV. Biased Language -Wiki

Model Training time (seconds)

LDA 334.6

NMF 3.6

II. News Website Classifie

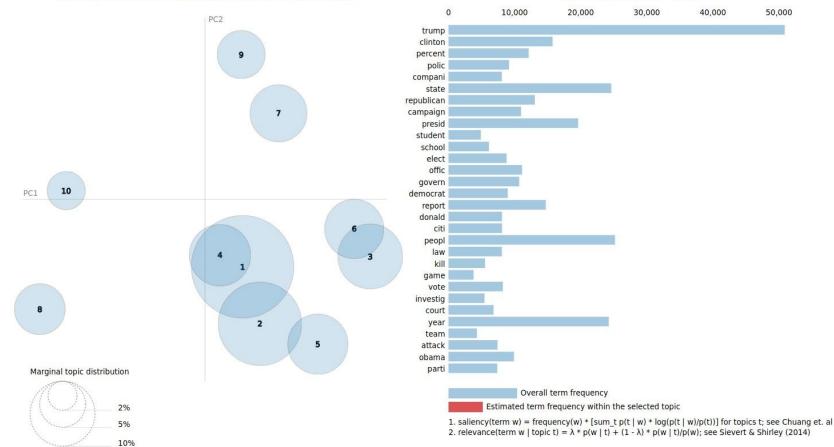
III. News Topic Modeling

IV. Biased Language -Wiki

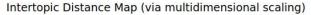
IV. Future



Top-30 Most Salient Terms1

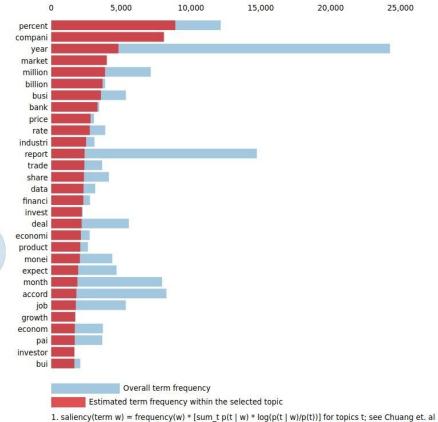


III. News Topic **Modeling**





Top-30 Most Relevant Terms for Topic 7 (7.7% of tokens)



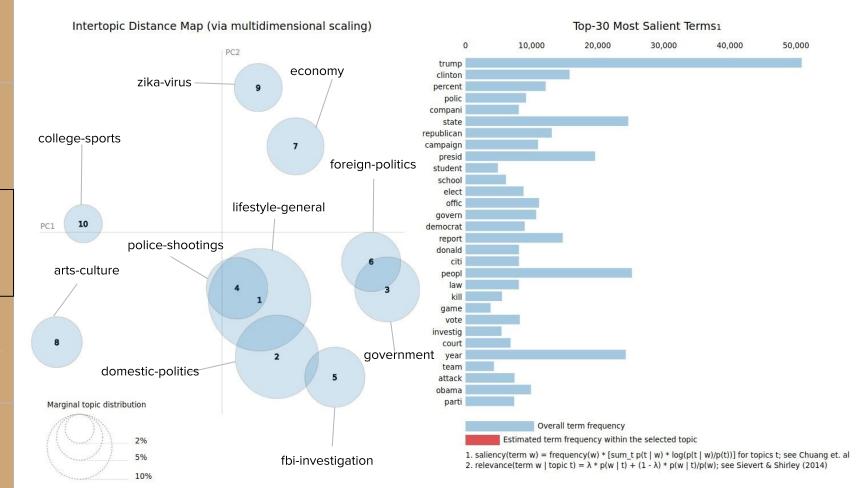
- 2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

II. News Website Classifier

III. News Topic Modeling

IV. Biased Language -

IV. Future



II. News Website Classifier

III. News Topic Modeling

IV. Biased Language Wiki

IV. Future A

Top 3 News Topics by Political Lean

Least Biased	
economy	39.7%
foreign-politics	19.7%
domestic-politics	9.4%

Left	
domestic-politics	25.8%
lifestyle-general	21.5%
police-shootings	10.8

Left-Center	
lifestyle-general	33.9%
domestic-politics	16.65%
government	8.7%

Right-Center	
lifestyle-general	32.8%
domestic-politics	16.65%
government	8.7%

Right	
domestic-politics	31.0%
lifestyle-general	26.1%
government	11.5%

Right-Extreme	
domestic-politics	37.1%
police-shootings	14.8%
lifestyle-general	14.0%

What is biased language?

II. News
Website

Examples [2]

"Shwekey's albums are arranged by many talented arrangers."

III. News Topic Modeling "The first research revealed that the Meditation technique produces a unique state fact."

IV. Biased Language - Wiki "Marriage is a holy union of individuals."

IV. Future . Questions

[2] - Pryzant et al., Automatically Neutralizing Subjective Bias in Text. AAAI 2020: 480-489

What is biased language?

II. News
Website
Classifier

Examples

"Shwekey's albums are arranged by many talented arrangers."

III. News Topic Modeling "The first research **revealed** that the Meditation technique produces a unique state fact."

IV. Biased Language - Wiki "Marriage is a **holy union** of individuals."

What is biased language?

II. News
Website
Classifier

Examples

"Shwekey's albums are arranged by many <u>different</u> arrangers."

III. News Topic Modeling "The first research <u>indicated</u> that the Meditation technique produces a unique state fact."

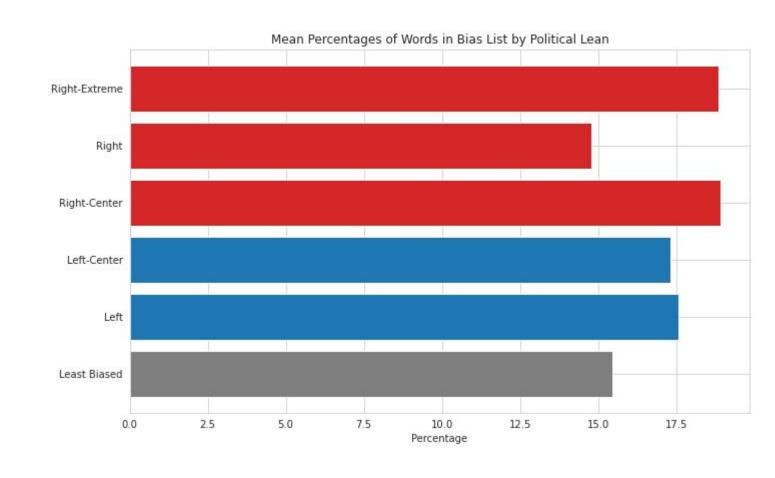
IV. Biased Language - Wiki "Marriage is a <u>personal union</u> of individuals."



II. News Website Classifier

III. News Topic Modeling

IV. Biased Language - Wiki



II. News Website Classifie

III. News
Topic
Modeling

IV. Biased Language -Wiki

IV. Future / Questions

Conclusions

- Both topic and type of language used tend to influence the degree of bias found in news media texts
- An NLP model that could accurately detect biased language would be useful for real-world editors of encyclopedias, neutral news sources, etc.

Future Work

 Conduct the same analysis on the corpus of Wikipedia articles that have been marked "Controversial"

Questions/Comments/Feedback