Variation in Political News

An NLP Approach

Stephen Lee

University of Memphis

October 22, 2019

Outline

Background

"If you can't measure it, you can't improve it."

— Lord Kelvin[1]

- The 2016 United States Presidential Election raised doubts for many about the quality of their news.
- Allcott and Gentzkow [2017] estimate that the average US adult read and remembered about one, and possibly up to several, fake news articles during the election period.
- While the "fake news" problem is often referenced, a full solution may not actually exist.

[1] The actual source of this quote is hard to pin down. While some attribute the quote to Lord Kelvin, others attribute it to Peter Drucker. A comment on stackoverflow further suggests that Antoine-Augustin Cournot was actually the first to express it in concise form in "De l'origine et des limites de la correspondance entre l'algebre et la geometrie" in 1847.

Literature Review: Computer Science

- Volkova et al. [2017], Wang [2017], and Shu et al. [2017] all look at detecting real vs. fake news.
- This isn't the same question, however. We need tools.
- The LSTM architecture was introduced in Hochreiter and Schmidhuber [1997]. Greff et al. [2016] show several variations to all be roughly equivalent.
- Schuster and Paliwal [1997] introduced the first bidirectional RNN.
- Together, bidirectional LSTM architectures prove to be among the most accurate models for language tasks, consistent with Wang et al. [2015].

Literature Review: Economics

- Gentzkow and Shapiro [2010] find that readers prefer to consume "like-minded" news and this can account for around 20% of the variation in political slant or bias.
- Gentzkow and Shapiro [2006] also show that a Bayesian consumer will reinforce their beliefs of a given news source quality when they read something that confirms their priors.
- Gentzkow and Shapiro [2008] suggest that competition in information markets may actually be counterproductive.
- Together, these works suggest that political news bias may tactical, and that this polarization we see may be self-reinforcing.

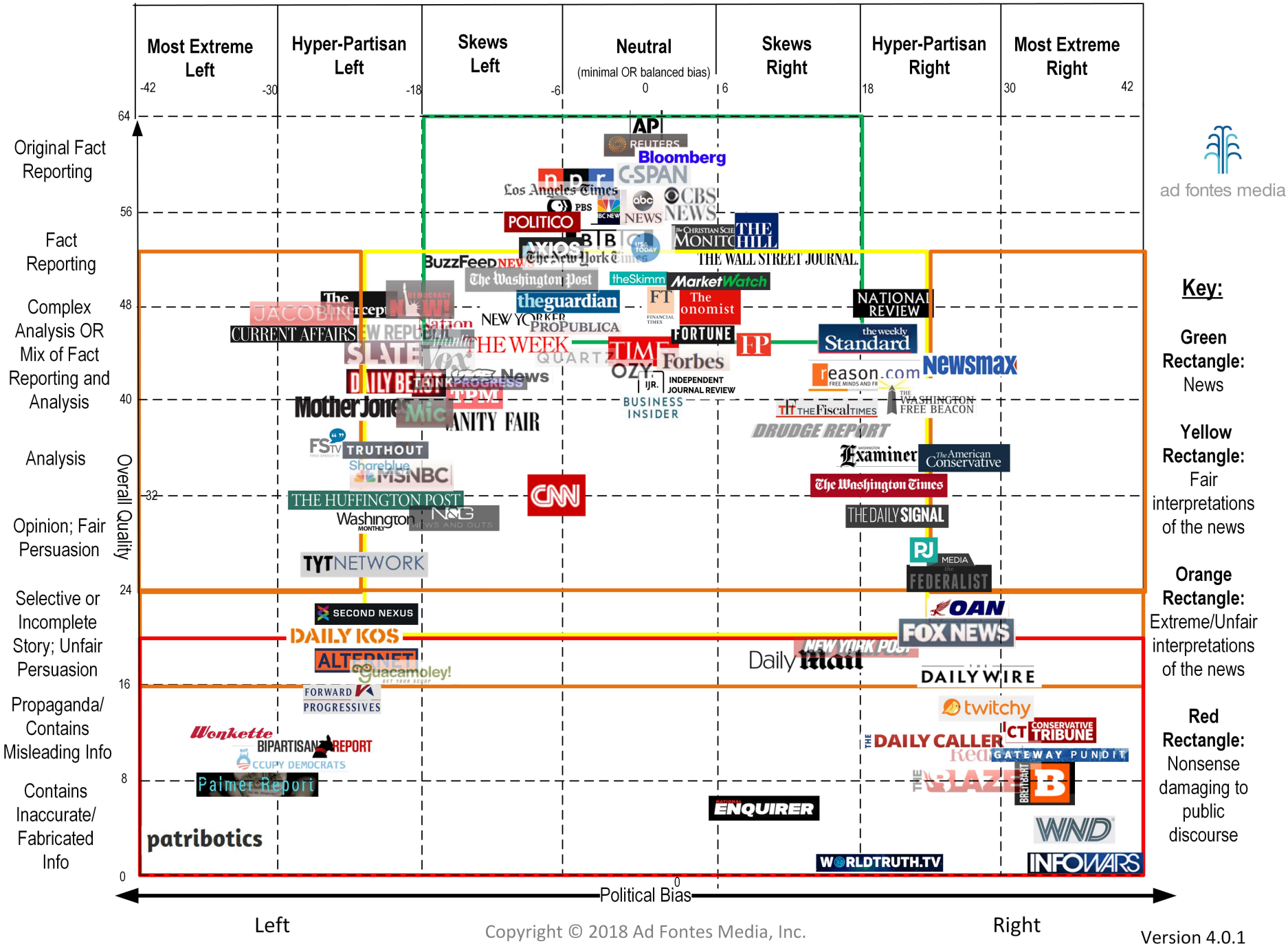
Summary

- With a goal of classifying articles by their news source, I scraped several thousand news articles from Fox, Vox, and PBS News.
- I found that a bidirectional, LSTM recurrent neural network can correctly predict the source of an article with high accuracy.
- I then used the result of this trained network to build a web app that can allow for a copy-and-paste interface to interact with this classification model.
- To my knowledge, this work provides the first thorough NLP based approach to understanding semantic differences across news sources.

Data

- I mined political news articles from the websites of Fox News, Vox News, and PBS News.
- Intuitively, the motivation is to focus only on sentiment or semantics, rather than subject matter differences.
- By some estimations,[2] these three news sites represent distinct categories of news:
  - Fox as a conservative right opinion.
  - PBS as the center primary source news position.
  - Vox as a liberal left opinion.

[2] Source: https://www.adfontesmedia.com.



ad fontes media

Key:

Green Rectangle: News

Yellow Rectangle: Fair interpretations of the news

Orange Rectangle: Extreme/Unfair interpretations of the news

Red Rectangle: Nonsense damaging to public discourse

Version 4.0.1
August 2018

Descriptive Statistics

Table: Summary statistics by news source.

Source	*Num Docs*	*Average word count*	*Pct adj*	*Pct adv*	*Average words / sentence*
Fox	661	686.2	6.6 %	3.4 %	20.1
PBS	1739	654.3	6.6 %	3.2 %	18.0
Vox	1027	1332.8	7.3 %	4.6 %	21.3

N-gram Frequencies

Table: Word Frequencies

Num	Vox	PBS	Fox
1	trump	trump	trump
2	tax	said	said
3	will	president	president
4	people	house	house
5	health	will	new
6	bill	new	will
7	republicans	white	democratic
8	one	senate	democrats
9	new	democrats	told
10	care	campaign	border

N-gram Frequencies

Table: Top frequencies of two word phrases.

Num	Vox	PBS	Fox
1	health care	white house	white house
2	white house	president donald	new york
3	trump administration	donald trump	president trump
4	donald trump	special counsel	green new
5	tax cuts	supreme court	health care
6	health insurance	attorney general	new deal
7	new york	new york	united states
8	affordable care	justice department	border security
9	tax bill	counsel robert	donald trump
10	federal government	trump said	state union

Challenges

1. Difference in corpus size from each source.
   - Bootstrap the data.
2. Variability of online formatting.
   - I removed any mention of their own organization.
   - Other common and unique affiliations.
   - Any other noticeable identifying characteristics.
3. Difference in the average article length.
   - I limited the article length to the first 500 words.

Statistical Analysis

- Based on Taddy [2013], I perform a statistical analysis of the data to see what words are most related to the various sources.
- Intuitively, consider speech as making draws from a multinomial distribution, based on your underlying "sentiment".
- Using the multinomial inverse regression, we can compare to see what phrases are most associated with which news source.

Vox and PBS

Table: Phrases that are indicitave of either Vox or PBS News.

	VOX		*PBS*	
1	email explain biggest	-6.90	chairman paul manafort	6.84
2	explain biggest news	-6.90	campaign chairman paul	6.84
3	biggest news health	-6.90	russia trump campaign	6.84
4	news health care	-6.90	mari clare jalonick	6.83
5	newslett check newslett	-6.90	trump campaign chairman	6.82
6	check newslett page	-6.90	mueller russia investig	6.82
7	mark email explain	-6.89	giant timelin everyth	6.81
8	health care edit	-6.89	timelin everyth russia	6.81
9	care edit sarah	-6.89	everyth russia trump	6.81
10	edit sarah kliff	-6.89	russia trump investig	6.81

PBS and Fox

Table: Phrases that are indicitave of either PBS or Fox News.

	PBS		*FOX*	
1	washington presid donald	-6.52	alexandria ocasiocortez dni	7.36
2	spoke condit anonym	-6.48	ongo partial feder	7.35
3	mari clare jalonick	-6.45	york democrat rep	7.35
4	timelin everyth russia	-6.44	alex pappa report	7.35
5	everyth russia trump	-6.44	chad pergram report	7.34
6	russia trump investig	-6.44	presid trump former	7.34
7	giant timelin everyth	-6.44	john robert report	7.34
8	read giant timelin	-6.44	adam shaw report	7.33
9	investig russian elect	-6.44	ap photoj scott	7.33
10	author speak publicli	-6.44	photoj scott applewhit	7.33

Vox and Fox

Table: Phrases that are indicitave of either Vox or Fox News.

	VOX		*FOX*	
1	email explain biggest	-6.45	nanci pelosi dcalif	7.43
2	explain biggest news	-6.45	partial feder govern	7.43
3	biggest news health	-6.45	kamala harri dcalif	7.43
4	news health care	-6.45	elizabeth warren dmass	7.42
5	newslett check newslett	-6.45	ongo partial feder	7.42
6	check newslett page	-6.45	york democrat rep	7.41
7	mark email explain	-6.44	greenhous ga emiss	7.41
8	health care edit	-6.44	major leader steni	7.40
9	care edit sarah	-6.44	leader steni hoyer	7.40
10	edit sarah kliff	-6.44	alex pappa report	7.40

Word Embedding

- I use the common crawl 840B Global Word Vector (i.e. GloVe).
- Introduced in Pennington et al. [2014] and uses 840 billion tokens and a case-sensitive vocabulary of 2.2 million words to map words into a corresponding 300×1 dimensional vector.
- Accordingly, I do minimal preprocessing to the text besides the basic cleaning mentioned previously.

LSTM Model

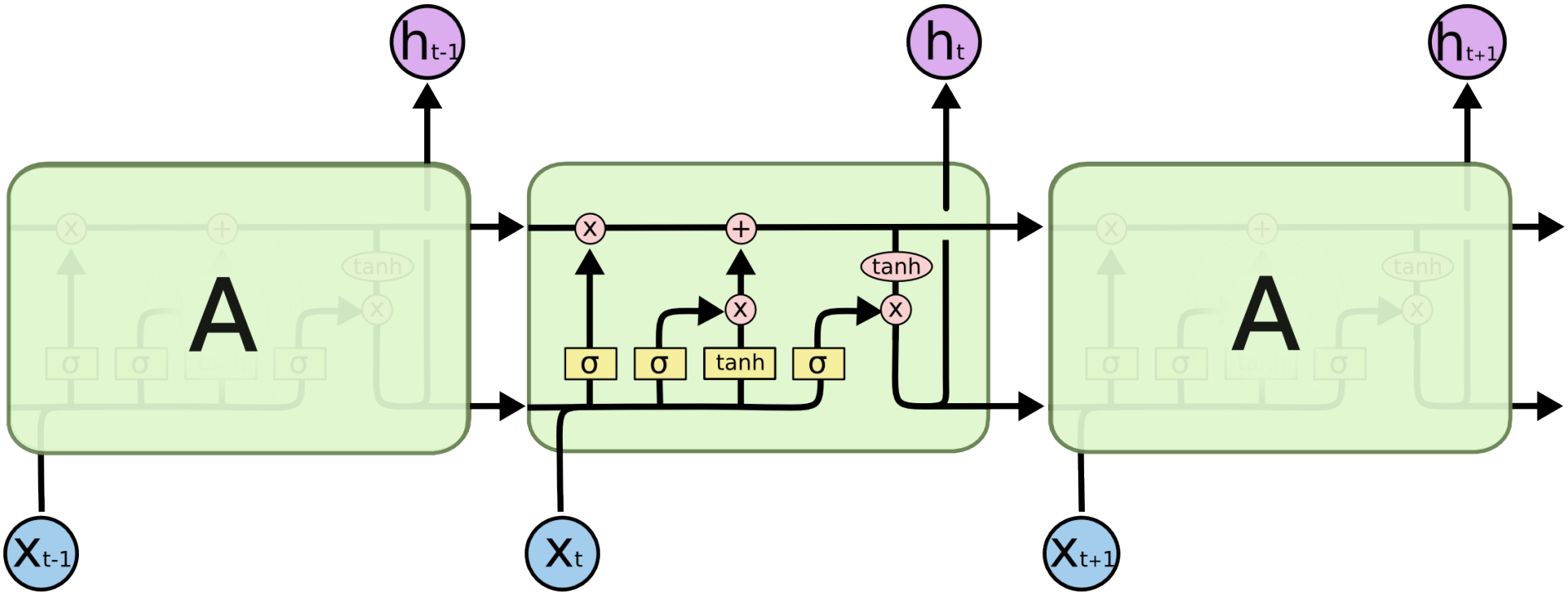


Figure: A graphical depiction of a single LSTM cell.[3]

[3] Source: https://colah.github.io/posts/2015-08-Understanding-LSTMs

Bidirectional Training

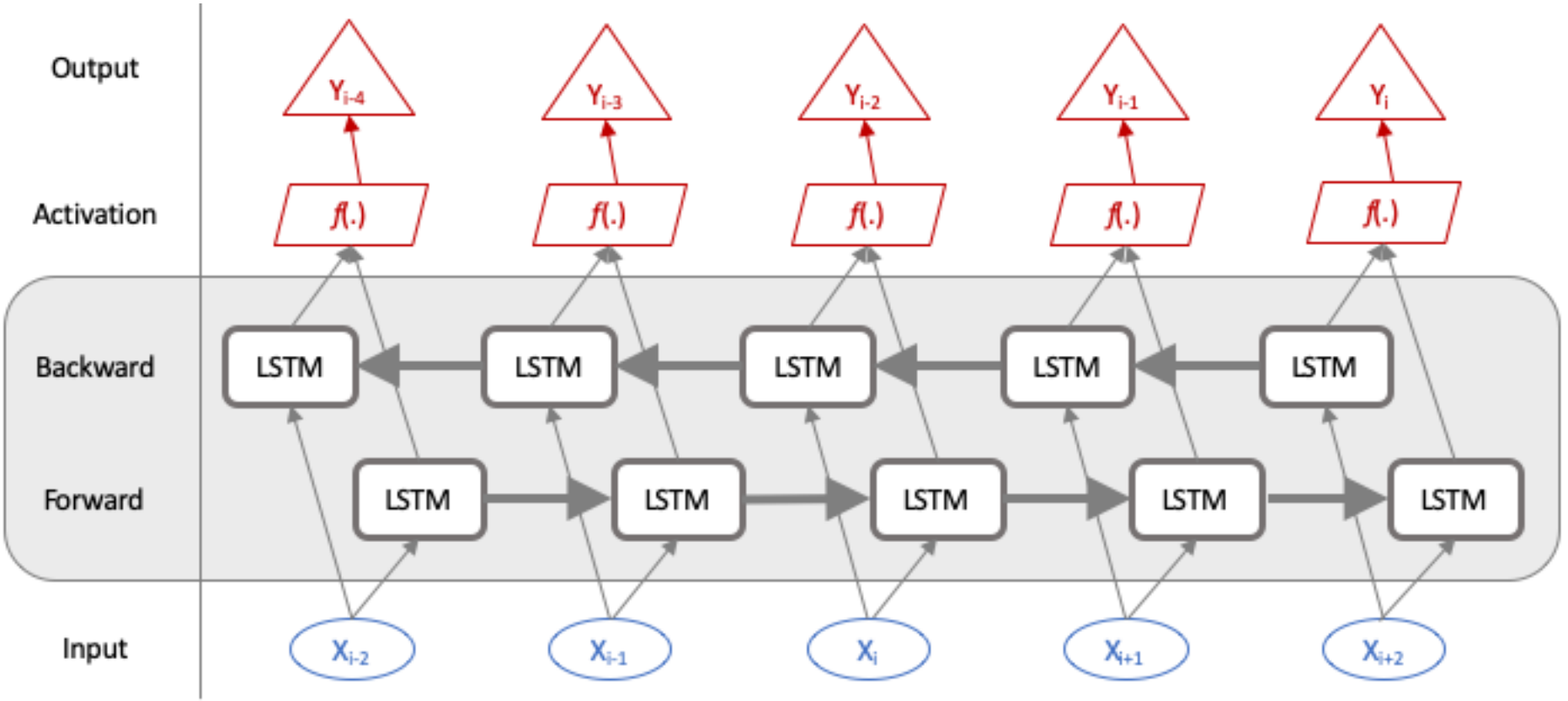


Figure: A visual representation of a bidirectional LSTM training.

Methodology

- I split the data into training (90%) and testing (10%).
- Training data: fit a birdirectional LSTM using a range of parameterizations.
- Testing data: make predictions, compare to actual, and calculate F1 scores.
- Similarly, I compare the bidirectional model to a forward only model using the same approach.

Note,

$$F_1 = 2 \frac{Precision \cdot Recall}{Precision + Recall}$$

Where,

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

Results: Bidirectional LSTM

Table: Training results from the bidirectional LSTM, sorted by F1 score.

Article Length	Batch Size	Dropout	Recurrent Dropout	Steps Per Epoch	F1
250	64	0.1	0.2	1000	0.946
500	64	0.2	0.2	1000	0.944
500	64	0.2	0.1	1000	0.939
250	64	0.1	0.1	1000	0.937
500	64	0.1	0.1	1000	0.937
500	64	0.1	0.2	1000	0.933
250	64	0.2	0.1	1000	0.921
250	32	0.2	0.1	1000	0.910
250	32	0.1	0.1	1000	0.906
...	...	...	. ..	...	...
500	32	0.1	0.1	500	0.828

Results: Forward LSTM

Table: Training results from the unidirectional LSTM, sorted by F1 score.

Article Length	Batch Size	Dropout	Recurrent Dropout	Steps Per Epoch	F1
250	64	0.2	0.1	1000	0.824
250	64	0.1	0.2	1000	0.797
250	32	0.1	0.2	1000	0.766
500	64	0.1	0.1	1000	0.724
250	64	0.2	0.2	1000	0.716
250	32	0.2	0.1	1000	0.703
500	64	0.2	0.2	1000	0.703
250	32	0.2	0.2	1000	0.695
250	64	0.1	0.2	500	0.686
...	...	...	...	...	...
500	64	0.2	0.1	500	0.556

Demo

Figure: The web app's home page. Link to page.

Conclusion and Discussion

- I've shown that NLP techniques can accurately classify news articles based on language differences in the underlying news sources.
- Given Gentzkow and Shapiro [2008] and Gentzkow and Shapiro [2006], it seems unlikely that biased news (or even fake news) will disappear anytime soon.
- A web application based on these ideas could serve as a starting point to measure bias in our news consumption. Analogous to:
  - Calendars can help to measure our time use.
  - Nutrition apps measure our macronutrients.
  - GPS measures our geographical position.
  - Can we have an app to measure our news consumption?

End

Thank you for your time.
Questions?

References

Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2): 211–36, 2017.

Matthew Gentzkow and Jesse M Shapiro. Media bias and reputation. *Journal of political Economy*, 114(2):280–316, 2006.

Matthew Gentzkow and Jesse M Shapiro. Competition and truth in the market for news. *Journal of Economic perspectives*, 22 (2):133–154, 2008.

Matthew Gentzkow and Jesse M Shapiro. What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1): 35–71, 2010.

Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10):2222–2232, 2016.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11): 2673–2681, 1997.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1):22–36, 2017.

Matt Taddy. Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, 108(503):755–770, 2013.

Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 647–653, 2017.

Peilu Wang, Yao Qian, Frank K Soong, Lei He, and Hai Zhao. A unified tagging solution: Bidirectional lstm recurrent neural network with word embedding. *arXiv preprint arXiv:1511.00215*, 2015.

William Yang Wang. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*, 2017.