AI For Everyone - Final Report

Bias Detection in Media Sources

: **Ruthran Chandrasekar, Ansul Shah, Caleb Getahun**

May 02, 2020

Abstract

The goal of our project is to compare the various biases present in left wing, right wing, and centrist media sources. We plan to create a separate language model for each of the three types of news sources, and run different tests to understand the bias in each model. Using the word2vec GenSim library along with leveraging transformers from the GPT-2 library, we hope to paint a vivid picture of the level of bias present in each model.

Key Links:

- Final Word2Vec Python Notebook
- Final GPT-2 Python Notebook

1 PROJECT DESCRIPTION

The goal of our project is to compare the various biases present in **left wing, right wing, and centrist media sources**. We plan to create a separate language model for each of the three types of news sources, and run different tests to understand the bias in each model. Using the **word2vec GenSim** library along with leveraging **transformers from the GPT-2 library**, we hope to paint a vivid picture of the level of bias present in each model.

To be precise, we hope to answer the following question: Is there a clear bias in different models created with varying political ideologies? If so, to what extent is each model biased?

We hypothesize that, if we are able to construct reliable, representative models of the various news sources, we should be able to find certain positive and negative associations of words based on the model we're observing. For example, inputting the phrase "President Obama" for the leftist model should give us positive or neutral word associations, but the same input into the rightist model will likely give us conservative-biased ideas, such as "Muslim", "weak", or "foreigner."

The impact of our results is far-reaching, as many NLP models are trained on news sources that may contain significant political or social biases. As NLP models spread and their results are implemented in daily life, we must be careful to avoid any unwanted bias or stereotypes that may be learned during training.

2 BACKGROUND RESEARCH

As unfortunate as it is, many Americans visit one or two news sources regularly and absorb those opinions and perspectives as their own. As a result, it's important for us to consider the inherent bias present in a variety of news sources that Americans frequent. In this sense, the results of our project will be helpful in understanding the extent of the bias that we encounter in everyday life.

The most similar paper we've found on the topic comes from Doumit et al. from the University of Cincinnati.[2] Their method utilizes a novel process called latent Dirichlet allocation (LDA), which is a probabilistic

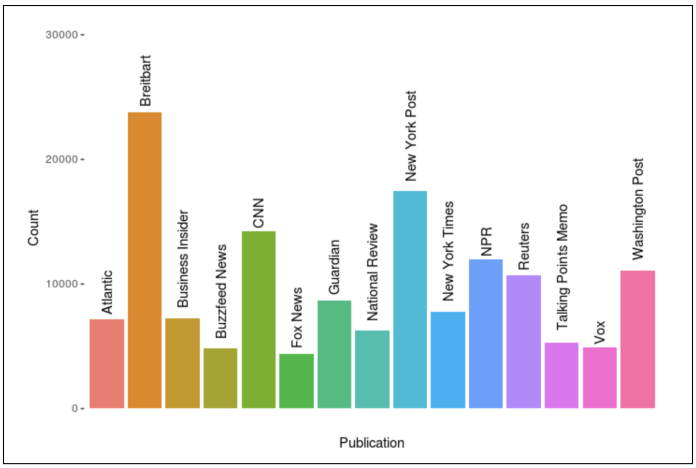

Figure 1. Article count distribution vs. news sources[1]

modelling tool in conjunction with NLP. The goal of the study is to analyze similarities in document structure combined with sentiment analysis to understand how different news outlets cover various news events. Their results focus on the top 10 associate words on news topics such as China, Wikileaks, and the Koreas from the NY Times, CNN, and other sources.

Our work is based on a similar idea of viewing news sources separately, but we are more focused on understanding the potential biases and stereotypes present in these news sources.

3 DATA DESCRIPTION

Our data includes **150,000** articles from **15** different news sources and contains information such as an articles title, date of publication, publisher, and article content. Of course, for our analysis, we'll be focusing on the article content to train our language models. The dataset can be found on **Kaggle**, called the "**All-the-news" dataset.**[1]

id	title	publication	author	date	year	month	url	content
17283	House Republicans Fret About Winning Their Hea...	New York Times	Carl Hulse	2016-12-31	2016.0	12.0	NaN	WASHINGTON — Congressional Republicans have...
17284	Rift Between Officers and Residents as Killing...	New York Times	Benjamin Mueller and Al Baker	2017-06-19	2017.0	6.0	NaN	After the bullet shells get counted, the blood...

Figure 2. The All-News dataset snippet[2]

4 PREPROCESSING STAGE

The preprocessing stage is vital for language models. Our approach required a few processes before training the model. The exploratory analytics were performed using the Pandas library in Python.

4.1 Clustering Sources

The first step after loading the data was to group the above news sources into leftist, rightist, and centrist news sources. We relied on a paper by Norregaard et al.[3] that examined misinformation in news articles. The research paper included a graphic that classified some of the most prominent news sources in America into far left, left, center, right, and far right sources, which we used to classify our own 11 sources into three categories. The classification of our sources is presented below:

Left:*Business Insider, Buzzfeed News, CNN, Guardian, New York Times, Talking Points Memo, Vox, Washington Post*

Right: *Fox News, New York Post, National Review, Breitbart*
Center: *NPR, Reuters, Atlantic*

4.2 Remove Punctuation

We want to remove the punctuation (such as commas, periods, semicolons, etc) because our model requires us to input a list of lists, where each list is a sentence. As a result, we need to remove these punctuation marks so that "president." is the same as "president" since the fact that the first version of the word is the last in a sentence shouldn't change its meaning or positioning in our **non-contextualised** W2V & D2V models.

4.3 Remove Stop-words

Since word2vec vectorizes a word based on the surrounding words, we don't want words such as "a", "of", and "the" to overpopulate the words and associate them with key words that carry meaning. Given a sentence such as "The president is a foreigner", a model will under-represent the connection between "president" and "foreigner" unless we remove the words that don't individually convey meaning to the sentence as a whole.

4.4 Word Lemmatizer

Word lemmatization is essential to our model not confusing different inflections of words since this could change the results our model gives back. We've extracted words using the nltk package that have similar inflections and grouped them as the root of the word so that there isn't a discrepancy. Words such as 'lead', 'leading', and 'led' do not show differences in our model since the root of the word will only be displayed.

5 MACHINE LEARNING METHODOLOGY

Since we want to create language models that allow us to test for bias by inputting various words into our three models to see the different word associations. Our methodology is split into two parts: **a)** static word embedding using word2vec models and **b)** eventually using contextualized word embedding through Finetune's library that utilizes transformers.

5.0.1 Mathematical Reference

$$Softmax(x)_i = \frac{exp(x_i)}{\sum_j exp(x_j))}$$

$$LogLikelihood = log(exp(x_i)) - log(\sum_j exp(x_j))$$

5.1 Word2Vec

We create word2vec models to analyze the word associations for different inputs. Because of the extremely long runtime to preprocess the data before learning a model, we sampled 12,000 articles from each cluster of news sources, meaning each of our three models was based on a random subset of 12,000 articles. In terms of the

model architecture, we used a continuous bag of words model (CBOW) with a window size (number of consecutive words) at 5.
Our choice to make a word2vec model was motivated by the simplicity of the model and the flexibility of finding the k-nearest words for any word input. Word2vec also uses very little memory, so it is preferable to other methods because it combines highly relevant functionality to our project while using very little time and memory.

5.2 GPT-2 Finetune Model

Leveraging the power of transformers was our primary goal from the outset of this project. The functionality of such a model, despite its significant learning time, is extremely useful for understanding trends and patterns in the text input.
As a result, we developed GPT-2 models from two different sources: a Buzzfeed News model to represent liberal views and a Fox News model to represent conservative ideas. Each of these sources have about 5,000 sources, which is a relatively small number of articles considering the size of our dataset. Despite this rather small input size, training time for each model took well over 3 hours on a GPU through Google Colab.
We decided to use transformers since they give a highly accurate representation of the dataset and can generate complex passages given a brief human input. This lends itself to an in-depth examination of bias, as we can produce a short article given a brief prompt and examine the different views on various subjects that each model generates.

6 FINAL RESULTS

***Note**: We have decided against the Doc2Vec model (which was proposed initially) in favor of a finetune GPT-2 model. So the content regarding any Doc2Vec work has been removed.*

6.1 Word2Vec Model

We're able to see some clear differences in how the different clusters of news sources cover different topics based on the common word associations.

Input word: **immigrant**

Leftist Model	**Centrist Model**	**Rightist Model**
undocumented	undocumented	**alien**
unauthorized	deport	undocumented
dreamer	illegal	illegal
muslim	citizen	deport
deport	muslim	foreigner
illegal	migrant	deportation
migrant	refugee	overstayed
latino	foreigner	citizenship
refugee	asylum	visa

Figure 3. We see the liberal sources with a slightly more neutral/positive coverage of immigrants, using words like "unauthorized" and "dreamer," whereas the rightist model is more inflammatory in its associations with the word "immigrant." The centrist model is pretty neutral in its associations

Input word: **clinton**

Leftist Model	Centrist Model	Rightist Model
crooked	**crooked**	**crooked**
sander	sander	frontrunning
trump	trump	**unlikeable**
delrayser	candidacy	sander
obama	throttled	surrogate
kaine	overeagerness	predation
romney	superdelegates	amplifies
bush		**lies**

Figure 4. Inputting the word "Clinton" gives us an idea of how the news cycle in general covered the Clinton campaign. Every model most highly associated the input with the word "crooked" (following the moniker "crooked hillary" pushed by President Trump)

6.2 **Finetune GPT-2**

Similar to the Word2Vec models, we provided the same input to the liberal model and the conservative model to see how each model completes the prompt. We get 10 model-generated passages from any given input, but many of the passages suffer from a lack of comprehension.
For example, given the prompt "Allowing more immigrants to come into the United States will...", the leftist model generated the following response that doesn't make logical sense:

> ...in fact, actually help the economy. For an average of $17, a day, immigrants in the United States make up just 10 percent of the workforce. That means that, on average, the average American will spend $2, a day, on the average immigrant, making up roughly one in nine American jobs.

Although the passage follows correct grammar and syntax rules, the actual content lacks clarity. Fortunately, by selecting the most realistic passages from the 10 outputs, we can get a clear idea of how each model views different topics.

Prompt: "Allowing more immigrants to come into the United States will..."

> **Left Model:** "...**boost job growth** to the point that it will have an impact on its economic activity. The **increase in employment** will also serve to **boost the economy**, especially in the manufacturing sector."
>
> "...not only **strengthen the economy**, but will also bring jobs that are hard to come by and **create millions of good paying jobs** every year."
>
> **Right Model:** "...not make America great. **It will only make America more dangerous**. The answer to all of Donald Trump's problems is a border wall. And it's not just a fence."

The left model includes a positive economic impact of immigrants, aligning with the liberal agenda that tends to be more inviting to immigrants. We found a couple different types of outputs from the right model: one which shows a negative portrayal of immigrants and another that emphasizes the economic benefits.

Prompt: "Anti gun violence protests are raging in many states."

> **Left Model:** "The NYPD is hosting a town hall event for gun owners that will air Wednesday night on WNYC Channel 4. Donald Trump has repeatedly said he would be "honored" to meet with Sen. Maria Cantwell of New York, a gun control advocate. Cantwell is one of several Democratic lawmakers who have been invited to a town hall event at 9:30 p.m. on Wednesday under the the National Rifle Association's (NRA) **"America's Problem With Gun Violence." "The NRA and their agenda of gun violence destroys families and destroys our country",** said the opening remarks of the event. **"They are out to get us."** One of the invitees, Rep. Chris Jackson of New York, said his district has "reasserted itself as the gun control advocate in the American dream." **"That's what I'm proud of."**
>
> **Right Model:** "The demonstration outside the headquarters of the National Rifle Association got underway shortly Thursday after the NRA's American Action Network delivered what it described as a "corner to the right." The gun advocacy group said, **"NRA leaders from across the country are joining forces with the #NeverHillary movement to ensure that Donald Trump does become president."** After years of standing up for rights, Trump on Thursday joined Democratic Hillary Clinton in urging gun control advocates to come together in the effort to secure a peaceful resolution to the nation's gun violence crisis. "We need you to come together in this effort," Trump said. **"Lock her up!"**

The models give us passages that align with their respective political ideologies, with the liberal model denouncing the NRA and its contributions to gun violence, while the rightist model combines two very conservative ideas, even adding in President Trump's classic campaign message: "Lock her up."

7 FUTURE WORK

As you can see, we've only developed an initial small scale 144MB GPT-2 instance which has many discrepancies in text generation. It allows us to focus on any contrast between the models, which works for our project's objective, but any further work on this would require a more sophisticated model developed using more data than current 4300 articles each, and a larger GPT-2 instance (1.5 GB). We are currently looking into the costs and availability of paid cloud services to issue these jobs and will possibly take this up during Summer Term I.

REFERENCES

1. All the news; 143,000 articles from 15 American publications, *https://www.kaggle.com/snapcrack/all-the-news* (accessed Mar 29, 2020)
2. Doumit, Sarjoun, and Ali Minai. "Online news media bias analysis using an LDA-NLP approach." In International Conference on Complex Systems. 2011.
3. Nørregaard, Jeppe, Benjamin D. Horne, and Sibel Adalı. "NELA-GT-2018: A large multi-labelled news dataset for the study of misinformation in news articles." In Proceedings of the International AAAI Conference on Web and Social Media, vol. 13, no. 01, pp. 630-638. 2019.
4. How to get started with Word2Vec — and then how to make it work, *https://www.freecodecamp.org/news/how-to-get-started-with-word2vec-and-then-how-to-make-it-work-d0a2fca9dad3/* (accesses Mar 29, 2020)
5. NLP for Beginners: Cleaning and Preprocessing Text Data, *https://towardsdatascience.com/nlp-for-beginners-cleaning-preprocessing-text-data-ae8e306bef0f* (accessed Mar 29, 2020)
6. Word2Vec Tutorial - The Skip-Gram Model, *http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/* (accessed Mar 29, 2020)