Detecting political bias in news media

Vivek Gupta
Srinivasa Arun Yeragudipati

Introduction

- The aim of our project was to detect news sources and political bias in various news media.

- We also tried to analyze whether it is possible to remove the bias by replacing certain words or phrases.

Methodology

- As discussed in the lectures, representation and structure are the two main aspects of any NLP model. We focused on the representation aspect of NLP.

- The reasoning behind using these representations was to see how well the bias detection model behaves in presence of contextualized sentence or paragraph embeddings

Experiments

There were two kinds of experiments we focused on in our project:

- Detecting the news source and political bias on media articles present in the datasets.

- Analysing the bias and devising a system to eliminate the bias from the text.

Datasets

Dataset	Task	N	c
Article Bias	Political Bias	37,554	3
NewB	News Source	264,000	11

Table 1: Comparison of the two datasets. N is the dataset size and c is the number of classes.

- The Newspaper Bias dataset (NewB) is a collection of over 200,000+ sentences regarding Donald Trump from eleven news sources.

- Article Bias Dataset consists of 37,554 news articles about various topics from various newspapers. Each of the documents is annotated as either left, centre or right.

Results - First Experiment

Features	Top-n Accuracy				
	n=1	n=2	n=3	n=4	n=5
word2Vec	0.262	0.407	0.513	0.601	0.684
sent2Vec	0.312	0.468	0.574	0.660	0.736
doc2Vec	0.252	0.401	0.514	0.605	0.685

Table 2: Top-n accuracies for each embedding on NewB dataset. Note that the dataset has 11 classes.

Results - Second Experiment

Sentence	O/T	L	C	N
he called trump a puppet a novice and an extremist	O	**0.53**	0.07	0.39
he called trump a puppet and a novice	T	**0.48**	0.08	0.44
he called trump a puppet	T	**0.52**	0.05	0.44
he called trump a genius	T	0.41	0.04	**0.55**

Table 4: Results from Experiment 2 with O representing the original sentence and T represents the transformed sentences. L here refers to Liberal, N refers to Neutral and C refers to Conservative

Discussion

- The project helped us gain valuable insights on different areas concerning the representation of language and structure modeling.

- If we had more time, we would actually try to create our own dataset by scraping articles from mostly the same news sources but cover different topics.