

Bias in Newspaper Headlines

Constantin Brîncoveanu

Timo van Nidek

Abstract—**TODO**

I. INTRODUCTION

Although newspapers tend to claim neutrality, it is almost impossible to find a source that does not contain any stereotypes or biases in some form. Biases in news headlines pose a danger in that they can propagate stereotypes to the general public. Detecting biases manually requires an enormous amount of effort to code news headlines and apply theoretical bias frameworks. It is infeasible to apply this process to the vast amounts of news stories that are released every day.

An alternative approach that has gained interest recently is to detect biases automatically using machine learning techniques. These systems employ natural language processing techniques to detect one of the many ways in which bias can be present in texts. One such method is sentiment analysis, in which the polarity of the text is determined, either positive, negative or neutral. Sentiment analysis is typically applied to highly subjective texts for which the polarity is very clear and has high variance across different texts. The difficulty with using sentiment analysis on news headlines to detect bias is that the polarity scores lie closely together, and therefore do not clearly indicate a bias on its own. News outlets report negative events such as war, death or protests more often than positive events, which makes it difficult to compare polarity scores. Additionally, bias detection using polarity as a measure is not enough since negative scores appear often, even unbiased reports. We therefore propose a technique that investigates the difference between the headline sentiment and the mean of the polarity scores for a specific topic.

Another method of investigating news bias is to measure the amount of flavor words used by a specific news source. These flavor words are adjectives, adverbs, comparatives and superlatives. The motivation behind this method is that these words can carry bias, and a truly objective report would need less of them. Our method compares the amount of flavor in a headline with the mean of the flavor across one topic to identify which news sources use significantly more flavor as an indicator of bias.

Our third and final approach is to classify news headlines into their respective sources, which will indicate how distinct a news source is. An unbiased news outlet would be very difficult to classify, whereas a news outlet that consistently uses biased language can easily be classified.

A. Definition of Bias

"Bias in cognitive science is generally defined as a deviation from a norm, deviation from some true or objective value." We examined deviations from the norm by calculating the mean topic sentiments, as well as the mean topic flavors, and then looking at the differences of the respective domains from those mean values.

A second way to detect biases was domain classification.

II. BACKGROUND

TODO

III. PROJECT

Our project consists of several approaches which let us detect biases and analyze newspapers from different perspectives. In this section, we are going to provide information about the dataset we used, as well as describe our approaches, ranging from general sentiment analysis, topic detection and sentiment and flavor analysis for the specific topics, to domain classification.

A. Data

Firstly, we looked at the Kaggle News Aggregator Dataset[1]. The advantage of this dataset is that it contains over 400k news stories from This dataset did not fulfill our requirements, because it only covered a quite small timespan in 2014 and it had a limited number of content categories, and no politic content.

Therefore, we started looking for alternative datasets, which led us to the discovery of the Reddit API. We selected newspaper headlines from "r/worldnews" ranging from 2016 to 2017. From those headlines, we further selected the ones that came from the most occurring domains.

B. Method

B.1 General Sentiment Analysis

One of our first results was a general Sentiment Analysis. We aggregated the results by averaging the sentiments of the headlines for each domain. Those average values ranged from -0.3 to -0.1. This revealed some interesting differences between the domains. It became clear that some domains had very negative average sentiments, because they focused on negative topics such as war and general world news, whereas others have relatively positive average sentiments, because they focused on economics or science news.

B.2 Topic Detection

As a more fine-grained measure of bias, we use sentiment analysis per topic. To extract the topics from our

data set, we use the biterm topic model (BTM) [2], which is especially suitable for short texts. BTM uses the word co-occurrence patterns (biterns) to create a model over the entire corpus, thus solving the problem of sparse word co-occurrences within one document. The model parameters are estimated using Gibbs sampling. We create a vocabulary of the top 5000 most occurring words in the corpus, and run the Gibbs sampling algorithm to model $T = 20$ topics. For the Dirichlet priors, we use $\alpha = 50/T = 2.5$ as recommended by [4] and $\beta = 0.01$ in order to make the topics contain a small number of highly distinct words. Due to limitations in computational power and the large size of the data set, we were limited to 250 iterations of Gibbs sampling, which took approximately nine hours to complete. We use a modified version of the BTM implementation found at [7].

B.3 Topic Sentiment Analysis

We determine the topic for all the headlines in the 2017 Reddit data set using the maximum topic probability. We extract the headlines for the selected topics in Table I and compute the polarity score for every headline. We use the VADER sentiment analysis framework [5] since it has been shown to achieve good performance on short social media text. VADER is able to detect the polarity of complicated sentences with human-like accuracy.

For every topic, we take the mean over the polarity scores of all headlines within the topic, which we call the *mean topic sentiment*. We then calculate the mean over the polarity scores within one topic for every news source separately. The difference between the mean sentiment per news source and the mean topic sentiment is an indicator of bias, using the definition of bias given in Section I.A.

B.4 Flavor Analysis

As an alternative metric, we introduce the concept of flavor. The amount of flavor in a headline is determined by the amount of words that can potentially carry subjectivity, namely adjectives, adverbs, superlative and comparatives. Intuitively, objective records of news events require less flavor words, while flavorful headlines are more likely to contain bias. We determine the flavor words using the part-of-speech tagger from NLTK [6]. Since longer sentences are more likely to contain flavor words, we define a normalized flavor metric as the number of flavor words, divided by the total number of words in a sentence.

Similar to the topic sentiment analysis, we compare the flavor per domain and topic with the average flavor taken over all headlines in one topic, the *mean topic flavor*. It is important to note that a deviation from the mean topic flavor does not indicate that a news source is negatively or positively biased for one topic. Instead, such a deviation means that a news source uses significantly more subjective words than others, which we consider an indicator of bias using the definition in Section I.A.

C. Evaluation

TODO

C.1 Topic Detection

Since there is no good quantitative measure to evaluate the quality of the detected topics, we instead validated them using a small-scale study in which three participants were asked to define a topic from the top twenty most distinct keywords as found by the BTM, and give a rating on a scale from 1 to 7 for how coherent the topic was. We selected a subset of the twenty topics based on average coherence rating and potential for controversy to perform sentiment analysis over. The selected topics can be found in Table I. After extraction of these selected topics, the data set was severely reduced in size. The distribution of headlines per topic is shown in Fig. 1, the number of headlines per news source after selecting topics is shown in Fig. 2.

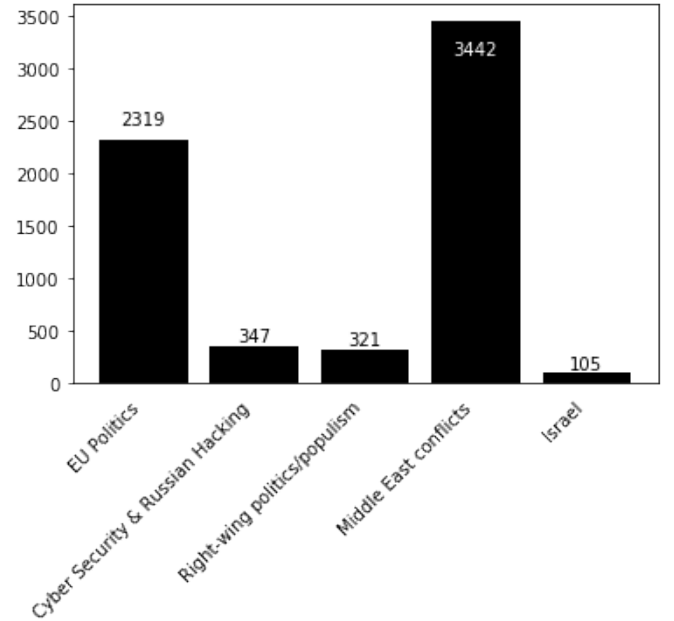


Fig. 1. Number of headlines per topic.

C.2 Sentiment Analysis

The deviation from the mean topic sentiment as a measure for bias was evaluated using a two-sided T-test.

C.3 Flavor

C.4 Domain Classification

We implemented a Bag of Words model with a Random Forest Classifier. For this task, we handpicked six domains. Those were the five most occurring domains: BBC, CNN, Independent, Reuters, and The Guardian. We additionally selected Foxnews as a sixth domain, hoping for it to be a contrast to the other five ones.

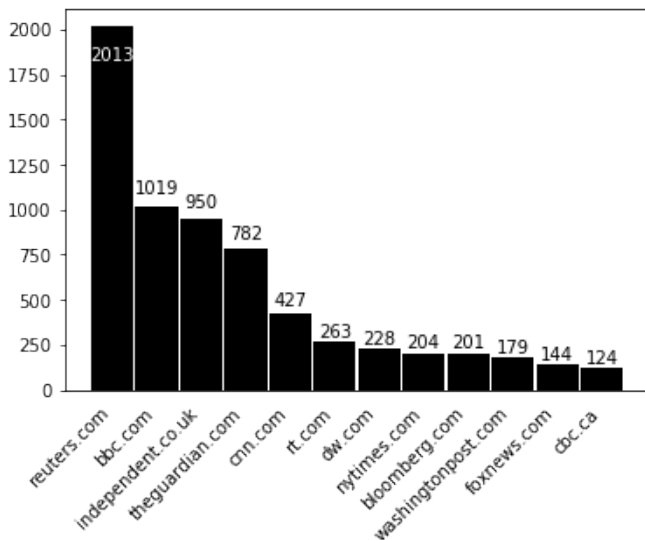


Fig. 2. Number of headlines per news source.

IV. RESULTS

A. Sentiment and Flavor Analysis

B. Domain Classification

V. SUMMARY

TODO

VI. CONCLUSIONS

There are measurable biases in newspaper headlines. Classification yielded surprisingly good results. The detected differences between the respective domains are statistically significant.

Sentiment Analysis is not accurate enough to detect the true sentiment value for each headline. In fact, the individual values often appear disappointingly inaccurate. However, if the number of headlines is big enough, the aggregated mean values per domain show significant differences from the general mean.

There are many more possible approaches to detecting newspaper biases, and we only scratched the surfaces of this broad topic. Further research is definitely going to reveal more.

REFERENCES

- [1] Lichman, M. (2013). *UCI Machine Learning Repository* <http://archive.ics.uci.edu/ml>. Irvine, CA: University of California, School of Information and Computer Science.
- [2] Xiaohui Yan, Jiafeng Guo, Yanyan Lan and Xueqi Cheng, “A Biterm Topic Model for Short Texts”, *Proceedings of the 22nd international conference on World Wide Web*, ACM, 2013, pages 1445–1456.
- [3] D. M. Blei, A. Y. Ng and M. I. Jordan, “Latent Dirichlet Allocation” *Journal of machine Learning research*, Jan. (2003), pp. 993–1022.
- [4] Thomas L. Griffiths and Mark Steyvers, “Finding scientific topics”, *Proceedings of the National academy of Sciences*, vol. 101, suppl 1, 2004 pages 5228–5235.
- [5] Clayton J. Hutto and Eric Gilbert, “Vader: A parsimonious rule-based model for sentiment analysis of social media text.”, *Eighth international AAAI conference on weblogs and social media*, Ann Arbor, MI, USA, June 2014.

- [6] Steven Bird, Ewan Klein and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*, “O’Reilly Media, Inc.”, 2009.
- [7] Joan Capdevila Pujol (Universitat Politècnica de Catalunya), “jcapde/Biterm: Biterm topic model”, *GitHub*, <https://github.com/jcapde/Biterm>, posted October 14, 2015, accessed May 24, 2017.

APPENDIX

TABLE I
SELECTED TOPICS WITHIN THE 2017 REDDIT DATA SET

Keywords	Topic	Average Coherence
election, EU, Brexit, vote, party, president, minister, UK, parliament, presidential, Theresa, May, says, prime, European, Le, Pen, government, would, French	EU Politics	5.67
Russian, Russia, US, intelligence, Trump, hacking, hackers, attack, cyber, CIA, FBI, security, government, former, ex, spy, data, says, officials	Cyber security & Russian hacking	6.00
Germany, party, Trump, right, says, German, Nazi, world, anti, social, pope, new, election, far, Brexit, leader, president, French, speech, saying	Right-wing politics & populism	4.33
Syria, Russia, US, north, Korea, Trump, military, missile, UN, war, Israel, turkey, Iran, state, said, Syrian, says, united, forces	Middle East conflicts	5
Israel, Trump, Israeli, president, Netanyahu, house, white, US, Palestinian, Donald, says, minister, Jerusalem, bank, peace, Palestinians, visit, settlements, west, two	Israel	5.66