

# Study of Detecting the Political Bias in News Articles

Pritesh Patil<sup>1</sup>, Shubham Chaudhari<sup>2</sup>, Arpita Dhote<sup>2</sup>, Mrunali Gorde<sup>2</sup>, Durvesh Palkar<sup>2</sup>

<sup>1</sup> Professor, Information Technology, AISSMS Institute of Information Technology, Pune, India

<sup>2</sup> Student, AISSMS Institute of Information Technology, Savitribai Phule Pune University, Pune, India

**Abstract :** Nowadays news audiences are experiencing a "echo chamber" due to news biased coverage, which causes individuals to shape views with only one side of the story in mind. Media is considered the strongest source of insight and viewpoint for readers about current events. Therefore, the inequality and misrepresentation of media portrayal of topics is necessary to scrutinise. So by computing a bias score for the news articles, the reader is able to make an informed decision related to particular events. In this paper, an attempt is made to prepare a survey report detailing some of the literature work done for figuring out the systems which detect the political bias for the news articles.

**Index terms :** *Media bias, News slant, Bias Detection, Automatic detection of media bias*

## I. INTRODUCTION

News coverage has a significant effect on the public as it characterises the incident as a whole and offers information on various topics such as climate, technologies and risk. News content, though, often shows an indoor bias, displayed in media articles and often seen as media bias or slanted news.

In the context of certain events and articles that are published and how they are covered, media bias is the bias or assumed bias by journalists and news publishers within the mass media. Ownership or journalists or sources of income of the media outlet, or a certain political point of view of the outlet and its viewers, are the reasons that may be blamed for media bias. Media bias has the effect of isolating the news viewer during a "political bubble" or "echo chamber," i.e. being influenced solely by news and views similar to their own.

Recently, in the field of computer science the study & automatic detection of media bias in news reports has gained interest. News aggregators are a common example, as Google News provides news users a quick summary of a wide spectrum of news. However, there is currently little assistance from existing frameworks to display the assorted viewpoints found in publications focusing on the same phenomenon. Hence by using a system which can compute the bias score for the news articles which directly provides news consumers with more options to make an informed judgement.

A few existing systems which give the bias score but some of them are not efficient while others are efficient but only on limited datasets, so our aim is to study all existing systems and evaluate a need of a system with less error & more efficient performance. The study discussed in this paper focuses on the methodologies implemented rather than the outcomes in the vast literature.

## II. LITERATURE REVIEW

### a) A Bias Aware News Recommendation System :

In this paper, they used the findings of a Pew Research analysis of readers of several major news media in the English language, placing the news channels on a scale from conservative to liberal, depending on the inclinations of their readers' political beliefs. The module receives news articles from a number of the same news outlets from the Pew report on the same issue as the document currently being read. After which, in the form of a graph with the

political orientation of the news source on the X axis and the bias score of the news article on the Y axis, the links of articles taken from the news outlets to the user are seen. The user therefore makes an informed choice about what kind of news article they want to read.

The system uses the pew report which could not be useful while browsing as well as not politically aware enough to judge how biased the articles were.

#### **b) Automated Identification of Media Bias by Word Choice and Labeling in News Articles :**

In this study, NewsWCL50, the WCL's first open appraisal dataset for bias recognition consists of 8,656 manual annotations in 50 news articles. Here is used the approach which is capable of removing instances of bias from WCL thus outperforming state-of-the-art approaches, such as coreference resolution, which currently can not address quite narrowly specified or abstract coreferences used by journalists. To test the process on the NewsWCL50 dataset, the highest performing state-of-the-art methodology achieves an F1=45.7 percent compared with F1=29.8 percent. Finally, they introduced a prototype showing the usefulness of the method in identifying frames affected by WCL bias.

The estimation of abstract or implicitly described frame properties is beyond the capabilities of current NLP as well as there is need to improve the estimation performance.

#### **c) Detecting Political Bias in News Articles Using Headline Attention :**

In this paper, a mechanism of headline coverage is used to automatically diagnose bias in news articles along with a manually annotated dataset to enable further analysis. The model generates a vector for news articles by aggregating essential words obtained by paying attention based on representation of headlines. The experimental findings indicate that the model outperforms all previous baseline models substantially. Attention mapping reveals how headline attention successfully sorts out bias-causing phrases. This model can also be applied to other sentiment-based text labelling such as blogs or trending internet posts featuring a title or headline and a frame.

There is a small difference in accuracy of predicted bias when comparing the headlines to considering the whole article.

#### **d) Political News Bias Detection using Machine Learning :**

This research uses the Multilayer Perceptron model, implemented by scikit-learn, Python's machine learning library. The MPL model consists of six models. The initialization module in MLP initialises and transfers the word embedding matrix to the training module that uses this matrix in the training dataset to vectorize terms & phrases. The resulting qualified classifier is invoked in the classifier module, which then takes the texts of the articles as its input and returns a percentage breakdown of the political ideologies. This output is then communicated to the Flask framework module, and sent for display to the Chrome extension. For the MLP model the word embedding matrix is initialised with the fastText word vector representation model from Facebook. The results of the classification task reveal that the MLP model outperforms the recurrent neural network (RNN) model with a appropriate collection of parameters on the IBC dataset of 4062 called sentences. Although the RNN received a 71.8 % F1 score, the MLP model was able to gain an 81 % F1 score.

Classification of Real-time Political News using the MLP classifier did not deliver optimistic outcomes.

**e) Political Ideology Detection Using Recurrent Neural Networks :**

Here, a recursive neural network (RNN) for detecting racial prejudice at sentencing is used. This model needs richer data than currently available, so researchers created a new dataset of political opinion annotated at the phrase stage. They demonstrate with this new dataset that RNNs not only identify phrases well but also enhance more when additional phrase-level annotations are provided. RNNs are quantitatively more productive than current approaches that independently use syntactic and semantic elements, and also demonstrate how social bias in complex syntactic constructions is correctly defined by the model.

Perhaps this model makes mistakes as polarity switches occur at nodes high up in the tree. RNNs are unable to model the longer IBC sentences effectively.

**f) Introduction to Multitask Ordinal Regression for Jointly Predicting the Trustworthiness and the Leading Political Ideology of News Media :**

In this research paper, they have focused on two under-explored problems which are “trustworthiness estimation” and “political ideology detection”. The system which has been proposed, models both the problems jointly using a multi-task regression framework. There is use of auxiliary tasks along with modelling centrality and hyperpartisanship and left vs right bias on a coarse-grained scale.

Methods used to solve those issues were corpora ordinal regression and auxiliary tasks. Those tasks were derived from the bias levels. The scale which was used was a 7 point biased scale. A sample of articles was being gathered from the target medium, and was being calculated features which are POS tags, linguistic cues, sentiment scores, complexity, morality, and embeddings. As a result it was observed that the single task of the ordinal regression model could perform more nicely than the majority class baseline which was based on both evaluation measures. It was further observed that the performance on the main task improves when jointly modeling several auxiliary tasks. Improvement was dependent on the auxiliary tasks which were in use.

The major drawback of this system is it only predicts the political bias, there are many more biases on which biased score could be calculated like the nationalist vs globalist, Islamic vs secular.

**g) Multi-view Models for Political Ideology Detection of News Articles :**

In this paper, a generalised MVDAM system model to exploit the statistical role of identifying news articles' political opinion from different perspectives. The integration of title, link framework and content defeats the state of the art by 10 percentage points on the F1 score effectively. Finally, using the model's expected probabilities, they use probability tuning approaches to identify news sources by their ideological proportions that are reasonably associated with recent studies on the ideological location of news sources.

**h) Towards Bias Inducing Word Detection by Linguistic Cue Analysis in News Articles :**

In this paper, the concept of inverse document frequency (IDF), which is a well-known statistic in the information retrieval, as one of the features to represent the word in the news article is used to detect bias inducing words in news articles. It presents a news bias detecting method in a hierarchical manner. The approach provided by this paper can be extended to other domains such as Wikipedia, as well as to newspapers where it is recommended to use unbiased language.

It does not have a labelled dataset to determine how effectively the method of identifying bias based on relative features captures bias in the newspaper by the use of human resources.

**i) Classifying the Political Leaning of News Articles and Users from User Votes :**

In this paper they have implemented three semi-supervised learning methods based on the classifications of political news articles and the users as conservative one or the liberal one which are on the basis of the assumption made that the liberal users will vote in the favour of liberal articles and conservative users in conservative articles. Few articles and users were labelled and the algorithms started tracing the political leaning labels to the complete graph. 96.6% accuracy was noticed for the user and 9.3% accuracy was noticed for the article by the best algorithm. There they have used manually coded labels that propagated to other people and articles. They used a naturally occurring source of data to propagate, a large file of descriptive “votes” for every article. They color-coded conservative one as red and liberal one as blue. They considered that the items that can’t be decisively labelled in the propagation process as “gray.” They used these three Learning algorithms: problem formulation, random walk with restart(RWR) Local Consistency Global Consistency(LCGC), absorbing random walk. As the conclusion, the best algorithm conquered 97.3% accuracy on articles and users to the decided target nodes.

The limitation of this case study is the ability of the algorithm can be improved. One more limitation of the propagation algorithm is that they require interactions between articles and users. As the unpopular articles were not covered by the site from which they used to get the data set which is Digg, their algorithm was unable to classify them.

**j) Mining Sentiment Classification from Political Web Logs :**

In this paper, researchers investigate the utility of Naïve Bayes and SVMs on a novel collection of datasets created from web log posts. The results show that the Naïve Bayes classifier significantly outperforms Support Vector Machines at a confident level of 99%, with a confidence interval of [1.425, 3.488]. Naive Bayes classifier is adaptive to the data set's class composition. A difference in the results is introduced by not having a fair class composition; the dominant class is more likely to be adequately categorised than the minority class. The bias effect due to the unbalanced nature of the datasets magnifies as the databases decrease in size.

It does not have the ability to track changes within people’s opinions on a particular topic and can not explore the time dependency of its data.

### **III. CONCLUSION**

The research and automated analysis of media biased in news articles has recently gained attention in the field of computer science. We addressed a literature analysis in this report on the identification of political ideology in news articles. The multiple and varied previous works, which cover many domains, are closely related and interdependent. By suggesting an easy method to spot political ideology in news articles, our report offers a detailed guide to analyze it.

## REFERENCES :

- 1]Patankar AA, Bose J, Khanna H. A Bias Aware News Recommendation System. arXiv preprint arXiv:1803.03428. 2018
- 2]Felix Hamborg, Anastasia Zhukova, Bela Gipp. Automated Identification of Media Bias by Word Choice and Labeling in News Articles. 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)
- 3]Rama Rohit Reddy, Suma Reddy, Radhika Mamidi. Detecting Political Bias in News Articles Using Headline. <https://www.aclweb.org/anthology/W19-4809.pdf>
- 4]Minh Vu. Political News Bias Detection using Machine Learning. <https://portfolios.cs.earlham.edu/wp-content/uploads/2018/12/senior-thesis-political.pdf>
- 5] Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), volume 1, pages 1113–1122
- 6] Ramy Baly, Georgi Karadzhov, Abdelrhman Saleh, James Glass, Preslav Nakov. Multi-Task Ordinal Regression for Jointly Predicting the Trustworthiness and the Leading Political Ideology of News Media
- 7]Vivek Kulkarni, Junting Ye, Steven Skiena, and William Yang Wang. 2018. Multi-view models for political ideology detection of news articles. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '18, pages 3518–3527, Brussels, Belgium.
- 8]Sora LIM , Adam JATOWT , and Masatoshi YOSHIKAWA. Towards Bias Inducing Word Detection by Linguistic Cue Analysis in News Articles. <https://db-event.jpjpn.org/deim2018/data/papers/275.pdf>
- 9]Daniel Xiaodan Zhou, Paul Resnick, and Qiaozhu Mei. 2011. Classifying the political leaning of news articles and users from user votes. In Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17- 21, 2011.
- 10]Kathleen T. Durant, Michael D. Smith. Mining Sentiment Classification from Political Web Logs.