Segregation of News Articles Based on the Political Bias

Pritesh Patil[1], Shubham Chaudhari[2], Arpita Dhote[3], Mrunali Gorde[4], Durvesh Palkar[5]

[1]*Professor, Information Technology, AISSMS Institute of Information Technology, Pune, Maharashtra, India*

[2,3,4,5]*Student, AISSMS Institute of Information Technology, Savitribai Phule Pune University, Pune, Maharashtra, India*

pritesh.patil@aissmsioit.org

cshubham173@gmail.com

arpitadhote235@gmail.com

mrunalgorde@gmail.com

palkardurvesh16032000@gmail.com

Abstract

Slanted or one-sided news coverage, which is widely referred to as media bias, may have a significant effect on individual readers and public perceptions about news subjects. People's inability to provide an intrinsic shield against their own prejudices has resulted in the dissemination of false facts, and is gradually taking us to a world where opinions are simply polarized. Modern innovation algorithms provide us with content that confirms our inherent biases and thus trapping us in our "echo chamber". As a result, in order to provide the end consumer with more options for making an educated decision, we agreed to introduce our suggested method of informing the public of how unjust the news article i.e., whether it has a liberal perspective or conservative perspective by calculating and displaying a bias score to compare the most recent article to others.

Keywords— **Media bias, News slant, NEWS crawler, RNN, LSTM**

I. INTRODUCTION

News articles are the prime source to gain the information about an event & thus play a crucial role in shaping individual and public opinion. However, media coverage often exhibits an invisible bias, reflected in news articles & commonly observed as media bias or slanted news.

Political polarization and fake or biased news have created a huge chaos in shaping the views of individual and public opinion related to events. Journalists, ownership, specific political point of view of the outlet and its viewers and source of income of that media outlet are all considerations that can lead to media bias. The single distributor of news can control what is shown to the news reader.

In order to consider every aspect of a multi-faceted incident, one must look at a variety of news outlets and piece their reports together to get a more complete picture of the observed situation. Unfortunately, most consumers have ignorance toward the opportunity or the desire to devote the time necessary to piece together their own interpretation of the tale before choosing which version, if any, to believe.

As a result, we devised a framework in order to provide many sides of multi-faceted incidents to news readers in order to make an informed decision that aimed to warn them about the bias present in that article as well as suggest other articles on the same subject from various news outlets.

II. RELATED WORK

The work proposed by Patankar AA, Bose J, Khanna H had used outcomes of a Pew Research survey[1]. The module receives news stories from a majority of the same news sources from the Pew survey on the same subject as the current text. The links of articles taken from news sources to the consumer are then shown in the form of a graph with the news source's political ideology on the X axis and the bias score of the news article on the Y axis. As a result, the user will make an educated decision on the kind of news report they choose to read. This paper focuses on the Pew Research analysis, which may not be helpful when browsing and may not be politically informed enough to determine how skewed the stories are.

The work proposed by Felix Hamborg, Anastasia Zhukova, Bela Gipp used the dataset NewsWCL50 for its transparent assessment[2]. It separates instances of bias from WCL, outperforming state-of-the-art methods such as coreference resolution, which can only resolve very specifically defined or abstract coreferences used by journalists at the moment. 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 demonstrated the method's utility in distinguishing frames influenced by WCL bias with a prototype. 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.

The work is proposed by Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, Philip Resnik[4] . Here, a model with a recursive neural network (RNN) is used for detecting racial prejudice at sentencing. This model needs richer data than currently available, so researchers created a new dataset of political opinion annotated at the phrase stage. A new dataset which is of political opinion annotated at the phrase stage is given as an input to this model. 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.

The work is proposed by Daniel Xiaodan Zhou, Paul Resnick, Qiaozhu Mei..Here, for classification of political news articles 3 semi-supervised learning methods are used [5]. An assumption has been made here that liberal users will cast their vote for liberal articles, and conservative users and articles will cast their vote for conservative articles. Lost in accuracy with the addition of datasets such as friendship links & HTML links where the correlation was lower & required interactions between stories and users.

Vivek Kulkarni, Junting Ye, Steven Skiena, and William Yang Wang, this model employs a MVDAM [6] framework that makes use of the statistical function of defining political leaning in news articles from multiple viewpoints. The integration of title, link framework and content defeat the state of the art by 10 percentage points on the F1 score effectively. Ultimately, they employ probability tuning methodologies to determine news sources based on their ideological proportions, which are reasonably associated with recent studies on the ideological location of news sources, using the model's expected probabilities.

III. SYSTEM OVERVIEW

The entire system is divided into three modules which are shown in Fig.1.

The three modules are:

• News Crawler
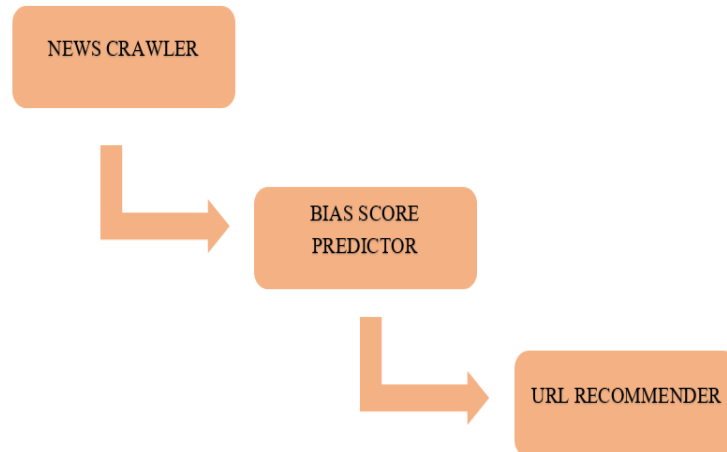
• Bias Score Predictor

• URL recommender

Fig. 1. Module split-up of system

There is a dependency of modules in this system. Without the successful execution of the previous module, the current module cannot be executed.
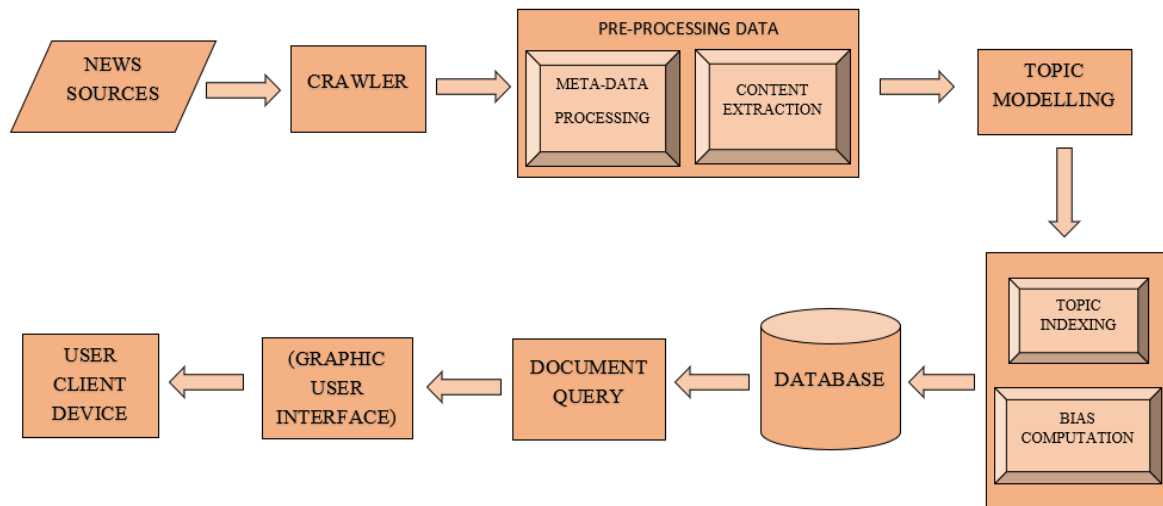

Fig. 2. Block Diagram of System

*A. Crawling news article*s

The News Crawler module does the job of crawling headlines from a wide range of conservative and liberal news sites, as well as mainstream and non-mainstream media organizations.

B. Prediction of Bias Score

The Bias Score Predictor module is used to obtain the bias score of the article. If the output value of Bias Score Predictor module is within the range [0, 0.5) then the article is of liberal perspective. And if the output value is within the range (0.5, 1], then the article is of a conservative perspective.
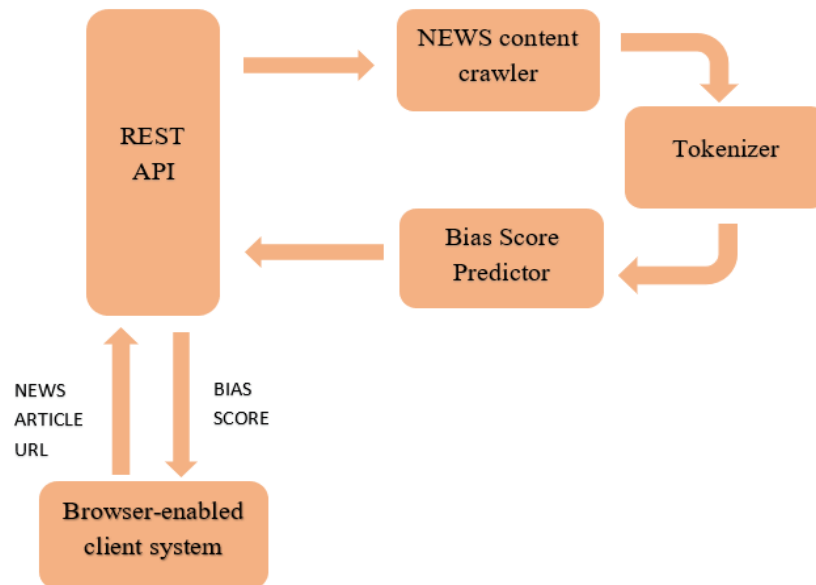

Fig. 3. High level Architecture of System

IV. METHODOLOGY

The first move in our program is to scrape news articles from a combination of outlets of conservative and liberal audiences, as well as mainstream and non-mainstream media sites. After that we are computing a bias score for the news articles & going to display bias scores for each article. After which we will recommend an option to see news articles from both liberal and conservative perspectives. You can see the entire workflow in Fig.2.To calculate bias score effectively we have used long short-term memory (LSTM) based Recurrent Neural Network (RNN).
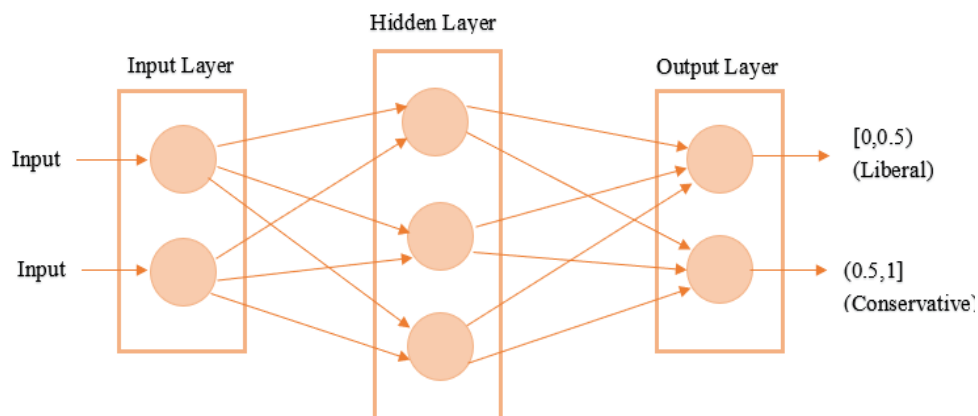

Fig. 4 RNN determines the bias score of article

The module for Segregation of News Articles based on political bias was accomplished by creating a website with a search bar as shown in the fig 5.

When a news topic is typed in the search bar and the "Search" button is clicked, runs a query through Google's search engine in an invisible window and looks at each URL rendered on Google's results page matching one of the trusted websites and then scrapes through the entire article. Next, the scraped text is then run through an RNN model trained using the IBC dataset. Here, for the classification purpose, we have used multiple-layered RNN. To classify the text, RNN convert each word in the article into a vector, which is then passed through the network and output has a scalar value which will be between 0 and 1, and if the output is less than 0.5, then the article is liberally-biased, and if the output is greater than 0.5 then article is conservatively-biased.

Segregation of news articles based on political bias

For what news do you want to search?

[   ] Search

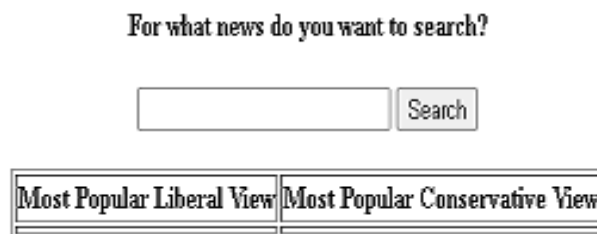[Most Popular Liberal View] [Most Popular Conservative View]

Fig. 5 Website

If the article fulfills a bias that is not already filled, then its URL is shown on the web-page. This process continues until the most relevant liberally-biased article and conservatively-biased articles are found, providing the user with a fair and balanced opportunity to see both political views (if either exists) about the news topic in question. Using the web-page, anyone can find a liberal and a conservative stance for any existing news topic being researched, ensuring users of the web-page access to a listing of relevant articles about a reported event that is statistically-proven to be fair and balanced. Moreover, when it comes to textual analysis, RNNs should be preferred over other models, particularly when it comes to sentiment analysis (e.g., political bias analyses). The importance of using LSTM nodes to mimic long-term memory has led to the RNN becoming one of the best, if not the best, machine learning algorithms to use for long-term contextual analysis.

V. RESULT AND DISCUSSION

In this study, we demonstrate how an RNN with LSTM nodes in one of its layers can achieve a respectable cross-validation accuracy of 58.26% and a high cross-validation F1 Score of 0.76 when evaluating political bias in news articles. The LSTM-based RNNs give a rewarding bias score. This combo can capture the long-term dependencies for example remembering the most frequently used words, that turn out fruitful in predicting the bias. Accuracies generated by standard vector machines (SVMs) and Naïve Bayes models are less useful for most textual analysis topics than RNNs have generated. However, since RNNs discover patterns, which provide knowledge including the context in which the most frequently encountered terms (keywords) are used. Iyyer et al.[4] evaluated bag-of-words models to RNNs for textual analysis and found RNNs to be significantly more reliable. We have successfully implemented all the three modules of our system. We have implemented the first module

i.e., NEWS crawler using python to collect news articles from various websites. Fig.6 shows how the NEWS crawler scrapes the news articles.

	website	url	headline	entities	Date
108	THE HINDU	https://www.thehindu.com/opinion/editorial/fall-from-grace-the-hindu-editorial-on-malaysian-ex-pm-najib-razak/article32225365.ece	Fall from grace: On Malaysian ex-PM Najib Razak	[(Malaysian, NORP)]	2020-07-30T00:02:00+05:30
109	THE HINDU	https://www.thehindu.com/opinion/editorial/digging-deeper-the-hindu-editorial-on-gst-compensation/article32215765.ece	Digging deeper: On GST compensation	[]	2020-07-29T00:02:00+05:30
110	THE HINDU	https://www.thehindu.com/opinion/editorial/the-cost-of-haste-the-hindu-editorial-on-drugs-vaccines-and-regulators/article32215774.ece	The cost of haste: On drugs, vaccines and regulators	[]	2020-07-29T00:02:00+05:30
111	THE HINDU	https://www.thehindu.com/opinion/editorial/in-reverse-gear/article32206858.ece	In reverse gear: on draft EIA notification	[(EIA, ORG)]	2020-07-28T00:02:00+05:30
112	THE HINDU	https://www.thehindu.com/opinion/editorial/a-governors-test-the-hindu-editorial-on-convening-rajasthan-assembly/article32198133.ece	A Governor's test: On convening Rajasthan Assembly	[(Rajasthan Assembly, ORG)]	2020-07-27T00:02:00+05:30
113	THE HINDU	https://www.thehindu.com/opinion/editorial/scandalising-as-contempt-the-hindu-editorial-on-proceedings-against-prashant-bhushan/article32198126.ece	Scandalising as contempt: On proceedings against Prashant Bhushan	[(Prashant Bhushan, PERSON)]	2020-07-27T00:02:00+05:30
114	THE HINDU	https://www.thehindu.com/opinion/editorial/arms-and-the-women-on-gender-barrier-in-indian-army/article32186249.ece	Arms and the women: On gender barrier in Indian Army	[(Arms, ORG), (Indian Army, ORG)]	2020-07-25T00:02:00+05:30
115	THE HINDU	https://www.thehindu.com/opinion/editorial/judicial-indiscipline-on-rajasthan-political-crisis/article32186245.ece	Judicial indiscipline: On Rajasthan political crisis	[(Rajasthan, NORP)]	2020-07-25T00:02:00+05:30
116	THE HINDU	https://www.thehindu.com/opinion/editorial/calibrated-balance-the-hindu-editorial-on-india-and-non-alignment/article32175684.ece	Calibrated balance: On India and non-alignment	[(India, GPE)]	2020-07-24T00:15:00+05:30
117	THE HINDU	https://www.thehindu.com/opinion/editorial/escalating-tensions-the-hindu-editorial-on-us-closing-chinese-consulate-in-houston/article32175704.ece	Escalating tensions: On U.S. closing Chinese mission	[(U.S., GPE), (Chinese, NORP)]	2020-07-24T00:15:00+05:30
118	THE HINDU	https://www.thehindu.com/opinion/editorial/another-front/article32165636.ece	Another front: On India, Bhutan and China	[(India, GPE), (Bhutan, GPE), (China, GPE)]	2020-07-23T00:02:00+05:30
119	THE HINDU	https://www.thehindu.com/opinion/editorial/justice-slow-but-sure-the-hindu-editorial-on-rajasthan-ex-royal-raja-man-singh-murder-case/article32165635.ece	Justice, slow but sure: On Rajasthan ex-royal murder case	[(Rajasthan, PRODUCT)]	2020-07-23T00:02:00+05:30
120	THE HINDU	https://www.thehindu.com/opinion/editorial/on-the-loose-on-west-bengal-governor-jaideep-dhankar/article32154141.ece	On the loose: On West Bengal Governor Jaideep Dhankar	[(West Bengal, GPE), (Jaideep Dhankar, PERSON)]	2020-07-22T00:02:00+05:30
121	THE HINDU	https://www.thehindu.com/opinion/editorial/cautious-optimism-on-race-for-vaccines-gather-pace/article32154144.ece	Cautious optimism: On the race for vaccines	[]	2020-07-22T00:02:00+05:30
122	THE HINDU	https://www.thehindu.com/opinion/editorial/no-right-answer-on-decision-of-reopening-of-schools/article32142676.ece	No right answer: On decision of reopening of schools	[]	2020-07-21T00:02:00+05:30

Fig 6. NEWS crawler

The statistical calculation of a neural network and the justification for its architecture have also been discussed; the properties of neural networks, as well as a description of what these algorithms actually produce, have also been given in this way. The reader types a news topic which he/she wishes to read in the search bar. For example, if the reader wants to search the term "Gun control" as shown in fig 7.

Fig 7. Searching for Gun control news articles

Based on the relevancy of the queried URLs to the reader's query, the site will find the first response on Google's search results. If the first article is conservatively tilted then RNN is listing it in the website's table's "Most Popular Conservative View" section. For liberally tilted posts, the procedure is the same. The links to the most relevant conservative article and the most relevant liberal article are

then shown in respective website's table. Webdriver.py is in charge of this task and it takes place in an invisible browser's Google search engine as shown in fig 8.
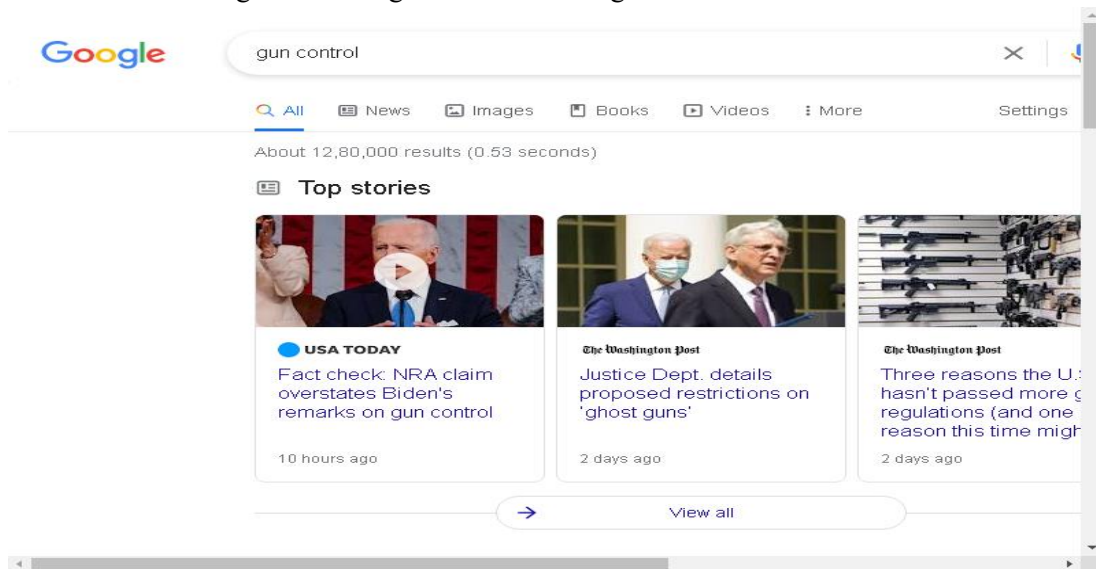

Fig 8. Invisible browser

Webdriver.py then waits for the "most relevant" URLs to load on Google's results page before reading and opening the first approved URL and running the articles' text via the RNN model developed. In order to distinguishes the text, conversion of each word in the headline of the article into a vector takes place by the RNN, which then passed through the network and gives the output as a scalar number as shown in fig 4.

Fig 9. Conservative biased Article

The bias score of the article in fig. 9 is less than 0.5 that's why it is liberally-biased article and bias score of the article in fig. 10 is greater than 0.5 that's why it is conservatively-biased.

Fig 10. Liberal biased Article

In this way, background activities include locating the most prominent conservative and liberal news coverage on a given subject which is "Gun Control" takes places and then posting links of both perspectives on the website side by side as shown in fig 11.

Fig. 11 The end result of search

VI. CONCLUSIONS

Our system offers a compromise between keeping user autonomy and the need to warn the users of biased news & readers will be able to see news articles from both liberal and conservative perspectives. It recommended the articles of the same topic from news sources with a variety of political biases and enabled users to be more aware of bias in news articles they are currently reading. It provides a balance between retaining consumer control and warning the users of distorted or misleading data.

ACKNOWLEDGMENT

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]   Arkajyoti   Misra   &   Sanjib   Basak   Political   Bias Analysis.https://cs224d.stanford.edu/reports/MisraBasak.pdf

[4]   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

[5]   Daniel Xiaodan Zhou, Paul Resnick, Qiaozhu Mei .Classifying the Political Leaning of News Articles   and   Users   from   User Votes. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2 782/3293

[6]   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.