

# A Bias Aware News Recommendation System

Anish Patankar, Joy Bose, Harshit Khanna

Samsung R&D Institute India, Bangalore

{ anish.p, joy.bose, h.khanna }@samsung.com

**Abstract**— In this era of fake news and political polarization, it is desirable to have a system to enable users to access balanced news content. Current solutions focus on top down, server based approaches to decide whether a news article is fake or biased, and display only trusted news to the end users. In this paper, we follow a different approach to help the users make informed choices about which news they want to read, making users aware in real time of the bias in news articles they were browsing and recommending news articles from other sources on the same topic with different levels of bias. We use a recent Pew research report to collect news sources that readers with varying political inclinations prefer to read. We then scrape news articles on a variety of topics from these varied news sources. After this, we perform clustering to find similar topics of the articles, as well as calculate a bias score for each article. For a news article the user is currently reading, we display the bias score and also display other articles on the same topic, out of the previously collected articles, from different news sources. This we present to the user. This approach, we hope, would make it possible for users to access more balanced articles on given news topics. We present the implementation details of the system along with some preliminary results on news articles.

**Keywords**— *Fake News, Recommendation System, Bias Detection*

## I. INTRODUCTION

Political polarization and fake or biased news have been topics of interest in recent times [1]. A number of solutions have been proposed to detect and flag fake news. Many of these solutions seek to prioritize trusted news sources at the server level, for example efforts by Google to partner with fact checking networks [2]. Another is to poll users for trust in news sources, as in Facebook's approach [3]. One criticism of such top down approaches is that they restrict user choice and is tantamount to censorship [4], since the user should be able to decide what kind of news articles they wish to read. In this paper, we present an alternative solution that seeks to give more choice to the end user to make an informed decision.

Our system first informs the reader how much the news article is biased by calculating and displaying a bias score to benchmark the current article with others, similar to the approach by Patankar et. al [5, 6]. After this, it offers recommendations from other news sources on the same topic as the current news article.

In our solution, we use the results of a Pew Research study [7] of the readers of some major English language news sources, positioning the news sources a spectrum from conservative to liberal on the basis of inclinations of political views of their readers. Our system extracts news articles on

the same topic as the currently read article from a few of the same news sources from the Pew study. We then display links of the articles taken from the news sources to the user in the form of a graph with the political inclination of the news source on the X axis and the bias score of the news article on the Y axis. This, we hope, would help the user make an informed decision of what kind of news article they wish to read. Our system gives an opportunity to the reader to explore articles that are different from their common political preferences.

The rest of this paper is structured as follows: in the following section we survey related work in bias detection and making a bias aware news recommendation system. Section 3 contains the outlines of our method, along with some implementation specific details. Section 4 presents the results of a user study to analyze how actual users liked the system. Section 5 concludes the paper.

## II. RELATED WORK

There are a number of works related to detecting bias in articles. Many of them use detection of sentiment bearing words [9, 10] to identify if an article is biased. Recasens et. al. [11] used a method to detect biased language, using a dictionary of words used by Wikipedia editors to ensure articles conform to Neutral Point of View (NPOV) rules. There has been recent work [22-23] on bandit based real time clustering techniques for news articles.

A few university students developed a plugin called OpenMind [14, 15] to counter fake news, as part of a recent Yale university hackathon. It gives the users feedback on their political biases in their reading patterns, as well as warn users if the source of the news articles is suspected to be fake. However, as per the details available online, it does not give feedback to the user as to how much biased is the current article they are reading. The methodology they use to detect fake news is also unclear.

A similar function is also claimed by BS Detector [16], a Chrome extension that uses a curated list of unreliable news sources to flag online articles as being fake news or otherwise unreliable. This too does not detect and state how much biased the article currently being read is, it is based on reliability of the source of the article rather than the content. Also, they offer no alternative article to read for the same topic. Our system, on the other hand, is aimed at making the user aware of the bias in the current article as well as recommend other articles on the same topic from different news sources, so the user can make an informed decision on what they want to read.

### III. SYSTEM OVERVIEW

In this section, we describe the steps of our approach for bias aware news recommendations in detail.

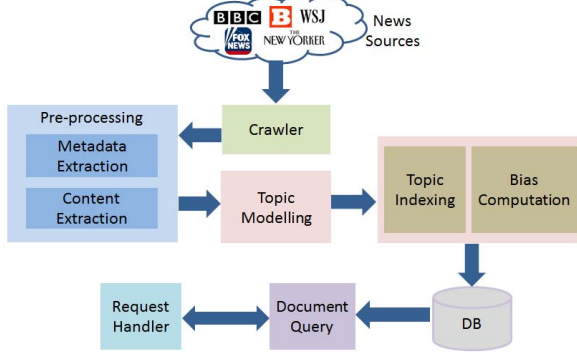


Fig. 1. Block diagram of the system for indexing news articles from different sources, along with their bias scores

#### A. Crawling news articles

The first step in our system is to crawl news content from a variety of news sources. The block diagram of the system is given in Fig. 1. We used selected news sources from the Pew Study [7], selecting a mix of sources with conservative and liberal readers, as well as mainstream and non-mainstream media outlets as per Wikipedia [8], where the top 6 media outlets are defined as mainstream. A graph of the scores of selected news articles from the Pew study is shown in Fig. 2.

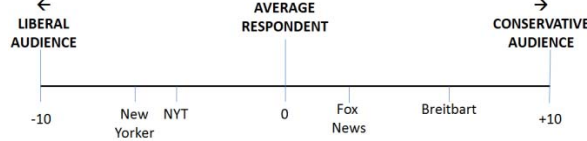


Fig. 2. Extract from the Pew Research study on the audience of selected news sources. Adapted from [4]

For each of the news sources selected, we crawl their websites for latest content, by fetching the robots.txt file from the new site domain URL, and then parsing the robots.txt file to extract the sitemap or sitemap index. After this, we parse the XML file of the sitemap and extract the URLs with timestamps. We then go through each of the URLs to extract the title and text content of the body of the webpage, as well as tags from the title if available.

Based on the availability of the robots.txt and sitemap files, we selected the following five media sources: New Yorker (liberal and non-mainstream), New York Times (liberal and mainstream), BBC (liberal and mainstream), Fox News (conservative and mainstream), and Breitbart (conservative and non-mainstream).

After collecting the text of the articles, we calculated and added the bias score for each article, as mentioned in the next subsection. We use this index to query similar articles from the crawled data.

For indexing, we used the semantic indexing system used by Sailesh et al [8]. The system uses an LDA based [13] indexing and matching system. For each article, the system

computes the topic distribution within the article. After that, it computes the cosine distance between two articles on the basis of the topic distributions, and on that basis builds a semantic association index.

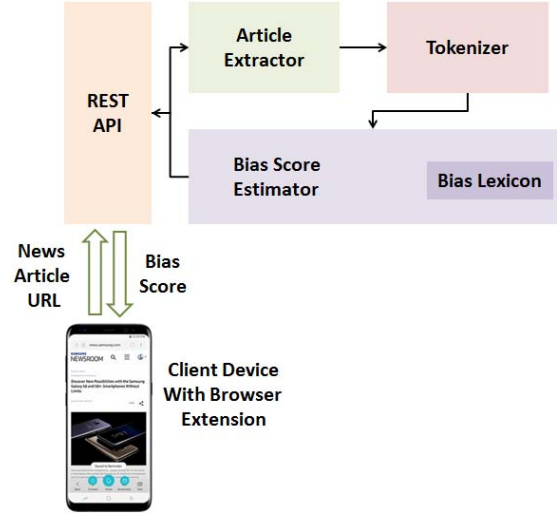


Fig. 3. High level architecture of the bias score computation system for one article.

#### B. Computing a bias score for the news articles

In this subsection, we describe the system for computing a bias score for an online news article. Bias is a subjective value, and hence any computed bias score is bound to be subjective too. However, we have validated our computed score with a user study to measure if the users rank biases of news articles similar to how our method would rank them.

One way to detect bias in a news article can be to examine if the sentences in the article contain any biased words, where the biased words can be chosen from a suitable created dataset. The article topic can be determined and can be compared with a generally agreed unbiased source such as Wikipedia article on the same topic. We can further examine the context in which the biased words are used. Another can be to examine the semantic meaning of the sentences and check if the chosen language used in the sentences is similar to the language used in other articles labelled as biased by users or from news sources that are considered biased or fake news. Another approach can be to perform part of speech (POS) tagging for the news article, and then measure how many of the adjectives used in the article are of the superlative kind.

In our approach, we use a bias detection system similar to the first one of the above three approaches (Patankar et. al [5, 6]). The system uses the non-NPOV lexicon from Wikipedia as created by Recasens [11]. Fig. 3 shows the architecture of the bias detection system for a given news article. The article bias score is found by totaling the biased sentences (containing one or more words from the bias lexicon) and dividing by the total number of sentences in the news article

Table 1 shows the algorithm to compute the bias score for a given sentence. We compute the aggregate similarity of the words in the sentence with words taken from a bias lexicon [11] taken from Wikipedia NPOV (neutral point of view) corpus. The NPOV corpus is used by Wikipedia

editors to determine which words to replace in an article to make it more neutral. Hence we chose this NPOV lexicon to decide which sentences had more words similar to neutral words, and use this to compute the bias score. In our approach, we computed similarity between words by using a Word2Vec model [17, 18] trained on Wikipedia English full articles corpus. Word2Vec creates word embedding vectors in such a way that the relations between words is preserved. As per the CBOW model, given a sequence of words (context) appearing together in the dataset, Word2Vec learns to maximize the following log likelihood for all words  $W_i$  in the device Id vocabulary:

$$\log P(W_i|\text{context}) = \frac{\text{similarity}(W_i|\text{context})}{\sum_k \text{similarity}(W_k|\text{context})} \quad (1)$$

After computing the bias score for each sentence in the article, the aggregate bias score for the news article is computed by first tokenizing the article into sentences, then adding the bias scores for each sentence and dividing by the total number of words.

TABLE I. ALGORITHM TO COMPUTE BIAS SCORE

COMPUTE-BIAS-SCORE-SENTENCE() // Compute NPOV score for given sentence	
1	$s = \text{sentence}; // \text{Input}$
2	$\text{biased\_words} = \text{List of biased words from the NPOV lexicon}; // \text{Input}$
3	$w2v\_model = \text{Word2Vec Model trained on Wikipedia articles corpus}; // \text{Input}$
4	$\text{count} = \text{count of similar words in sentence}; // \text{Input}$
5	$\text{similarity\_threshold} = \text{Threshold for similarity}; // \text{Input}$
6	$\text{npov\_count} = \text{NPOV score for sentence}; // \text{Output}$
7	$\text{total\_words} = \text{GET-WORD-COUNT}(s); // \text{Word count in the sentence}$
8	$\text{npov\_count} = 0;$
9	$\text{min\_distance} = 1.0;$
10	<b>for</b> $i: 1$ <b>to</b> $\text{total\_words}$
11	<b>if</b> $(\text{word in } w2v\_model)$
12	<b>for</b> $\text{biasword in } \text{biased\_words}$
13	$\text{similarity\_score} = w2vmodel.similarity(\text{word}, \text{biasword}); // \text{Word2vec similarity score of the word with the words in the Wikipedia NPOV lexicon}$
14	<b>if</b> $\text{similarity\_score} > \text{similarity\_threshold}$
15	$\text{npov\_count} = \text{npov\_count} + 1;$
16	<b>break;</b>
17	<b>return</b> $\text{npov\_count};$

### C. RESTful service for bias aware news recommendations on the same topic

Our system has a web browser extension that acts as a client to the web service. It send the article URL to the service which returns the resulting bias score for the article as well as the URL recommendations.

On the server, the URL is fetched, the URL content is extracted and queried in our database for matching articles from different news sources.

On the client, the browser extension queries the RESTful service and show the results to the user and allows the user to choose the article.

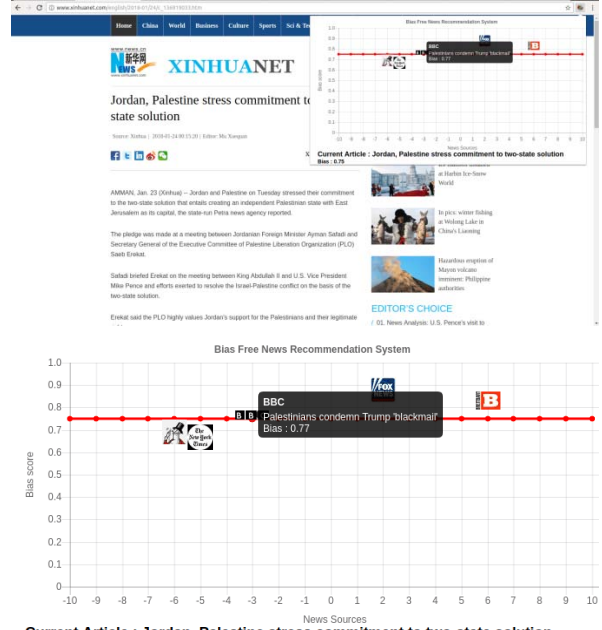


Fig. 4. Screenshots of the popup showing the related news articles from different sources on the same topic, on a desktop web browser for a news article.

### D. Implementation for a desktop browser

Fig. 4 gives a screenshot of the implementation for a desktop browser. On clicking the browser extension icon to the right of the address bar, the system displays a graph showing the current bias score on the Y axis and the news sources on the X axis. The current article bias score is shown in red, while the icons of the news sources represent links to news articles from those sources on the same topic. On hovering on the links, the user can see the news article topic along with the bias score.

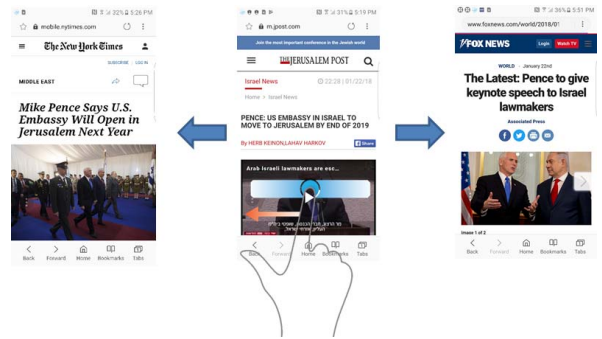


Fig. 5. Screenshots of the solution on a mobile phone, where the user has to swipe left or right with their finger to see articles from other news sources on the same topic, utilizing the semantics based browsing framework from [8].

### E. Implementation on a mobile phone

We also implemented the system on a web browser in the Android mobile phone, using the semantic browsing framework described in [8]. Here, instead of displaying the links along with the bias scores in a table, we displayed the

icons of the news sources on the same topic at the bottom of the web browser. The user flicks the news article left or right with the finger to see and read other articles from different news sources on the same topic. Fig. 5 shows the screenshots of related articles on the same topic for a mobile web browser.

In the following section, we present the results of some experiments to measure the bias score for news articles and validate our bias ranking system with a user study.

#### IV. EXPERIMENTAL SETUP AND RESULTS

We implemented the system on a Chrome web browser on the desktop with a browser extension, as well as on the default browser on a Samsung Galaxy S6 mobile phone.

##### A. Relative Bias of Different Types of Articles

To evaluate the bias scores generated by our model for news articles from different sources, we ran our tests on 10 Wikipedia featured articles and 10 Wikipedia non-featured articles. The featured articles are considered to be Wikipedia's best articles and often used by editors as examples. We also ran the tests for recent blog or opinion articles written by journalists in international media sources, as well as a few news articles on current topics. We determined the bias score for each of these articles.

The results are plotted in tables 2-5. We can see that the Wikipedia articles generate lower bias scores than non Wikipedia articles, and bias scores for media blogs (table 4) are typically more than Wikipedia (tables 2-3) and news articles (table 5).

TABLE II. BIAS RESULTS FOR WIKIPEDIA FEATURED ARTICLES

Wikipedia Featured Article Topic	Mean Bias Score
(a)Apollo 8	0.694
(b)Inocybe Saliceticola	0.663
(c)Thomcord	0.662
(d)Parity of Zero	0.749
(e)Denbies	0.663
(f)Lazarus Aaronson	0.713
(g)Eastbourne Manslaughter	0.724
(h)Daniel Lambert	0.670
(i)Oliver Typewriter Company	0.693
(j)Museum of Bad Art	0.735

Sources: (a) en.wikipedia.org/wiki/Apollo\_8  
(b) en.wikipedia.org/wiki/Inocybe\_saliceticola  
(c) en.wikipedia.org/wiki/Thomcord  
(d) en.wikipedia.org/wiki/Parity\_of\_zero  
(e) en.wikipedia.org/wiki/Denbies  
(f) en.wikipedia.org/wiki/Lazarus\_Aaronson  
(g) en.wikipedia.org/wiki/Eastbourne\_manslaughter  
(h) en.wikipedia.org/wiki/Daniel\_Lambert  
(i) en.wikipedia.org/wiki/Oliver\_Typewriter\_Company  
(j) en.wikipedia.org/wiki/Museum\_of\_Bad\_Art

TABLE III. BIAS RESULTS FOR WIKIPEDIA (NON-FEATURED) ARTICLES

Wikipedia Non-Featured Article Topic	Mean Bias Score
(a)Bernie Sanders	0.720
(b)Donald Trump	0.706
(c)Influenza A virus subtype H1N1	0.676
(d)Turing Machine	0.667
(e)John Glenn	0.673
(f)Royal Challengers Bangalore	0.633
(g)Dugong	0.740
(h)Purple Heart	0.676
(i)Blade Runner	0.701
(j)St. Petersburg	0.624

Sources: (a) en.wikipedia.org/wiki/Bernie\_sanders  
(b) en.wikipedia.org/wiki/Donald\_trump  
(c) en.wikipedia.org/wiki/Influenza\_A\_virus\_subtype\_H1N1  
(d) en.wikipedia.org/wiki/Turing\_machine  
(e) en.wikipedia.org/wiki/John\_Glenn  
(f) en.wikipedia.org/wiki/Royal\_Challengers\_Bangalore  
(g) en.wikipedia.org/wiki/Dugong  
(h) en.wikipedia.org/wiki/Purple\_Heart  
(i) en.wikipedia.org/wiki/Blade\_Runner  
(j) en.wikipedia.org/wiki/St\_petersburg

TABLE IV. BIAS RESULTS FOR MEDIA BLOGS

Blog Topic	Mean Bias Score
(a)E. Lewis, CNN: What's behind Trump's prayer in the oval office	0.806
(b)S. Larson, CNNTech: Malware researcher helps teen hackers turn skills into careers	0.840
(c)D. Ivory, NYT: The Deep Industry Ties of Trump's Deregulation Teams	0.809
(d)A. Carter, Washington Post: How to make the Islamic State's defeat last	0.830
(e)B. Majumdar, Economic Times: Victory in Champions Trophy to give new lease of life to Pakistan cricket	0.799
(f)M.S. Aiyar, NDTV: Reading Modi's Many, Fervent Hugs For Netanyahu	0.720

Sources: (a) edition.cnn.com/2017/07/13/opinions/pray-oval-office-trump-opinion-louis/index.html  
(b) money.cnn.com/2017/07/12/technology/malware-researcher-helps-teen-hackers/index.html  
(c) nytimes.com/2017/07/11/business/the-deep-industry-ties-of-trumps-deregulation-teams.html  
(d) washingtonpost.com/opinions/how-to-finally-defeat-the-islamic-state-and-make-it-last/2017/07/12/4d72ecc8-6717-11e7-a1d7-9a32c91c6f40\_story.html?utm\_term=.e7e8dc8f1550  
(e) blogs.economictimes.indiatimes.com/flavours-of-kolkata/victory-in-champions-trophy-to-give-new-lease-of-life-to-pakistan-cricket/

(f) [ndtv.com/opinion/reading-modis-many-fervent-hugs-for-netanyahu-1722985](http://ndtv.com/opinion/reading-modis-many-fervent-hugs-for-netanyahu-1722985)

TABLE V. BIAS RESULTS FOR NEWS ARTICLES

News Article Topic	Mean Bias Score
(a)E. McMurtry, Alternet: 6 Awesome Things About Bernie Sanders You Might Not Know	0.755
(b)Mises Institute: The economics of Bernie Sanders Review	0.794
(c)WSWS: Trump, the ugly reality of American politics	0.767
(d)Breitbart: Donald Trump vows to be the greatest jobs president God ever created	0.791
(e)TechRadar: iPhone6 Review: Growing old with grace	0.726

Sources: (a) [alternet.org/election-2016/6-awesome-things-about-bernie-sanders-you-might-not-know](http://alternet.org/election-2016/6-awesome-things-about-bernie-sanders-you-might-not-know)  
(b) [mises.org/library/economics-bernie-sanders](http://mises.org/library/economics-bernie-sanders)  
(c) [wsws.org/en/articles/2015/12/10/pers-d10.html](http://wsws.org/en/articles/2015/12/10/pers-d10.html)  
(d) [breitbart.com/big-government/2016/02/09/donald-trump-vows-to-be-the-greatest-jobs-president-god-ever-created/](http://breitbart.com/big-government/2016/02/09/donald-trump-vows-to-be-the-greatest-jobs-president-god-ever-created/)  
(e) [techradar.com/reviews/phones/mobile-phones/iphone-6-1264565/review](http://techradar.com/reviews/phones/mobile-phones/iphone-6-1264565/review)

These results (i.e. Wikipedia articles having a lower bias score) correspond to what we would expect, given that Wikipedia articles, being part of a peer edited encyclopedia, are supposed to be less biased or to use neutral language compared with news articles and media blogs.

Also, news articles have a different writing style and Wikipedia is more fact based rather than opinion based, so their lexicon is different. This could account for the difference in bias score.

#### B. Correlation of algorithm and user generated bias scores

We studied the bias scores obtained for the current article and recommended articles, for a variety of news articles on relevant current topics.

We conducted a small user study with 10 users to gauge (a) the accuracy of the system and (b) its usefulness and ease of use. We gave the users 3 sets of 5 articles each (each set of 5 articles being from the same topic but from different news sources), and asked them to rank the articles for the level of perceived bias. Table 6 gives the bias scores generated by our algorithm along with the URL recommendations (with bias scores) for the three sets of articles.

TABLE VI. BIAS SCORES ALONG WITH URL RECOMMENDATIONS FOR A FEW SAMPLE ARTICLES

Query URL and bias	Matching article URL	Bias Score
wsws.org/en/articles/2018/01/06/pakistan-j06.html Bias : 0.82	<a href="http://newyorker.com/news/news-desk/the-tapi-pipeline-and-paths-to-peace-in-afghanistan">newyorker.com/news/news-desk/the-tapi-pipeline-and-paths-to-peace-in-afghanistan</a>	0.72
	<a href="http://nytimes.com/2018/01/09/opinion/pakistan-trump-aid-engage.html?mtrref=undefined">nytimes.com/2018/01/09/opinion/pakistan-trump-aid-engage.html?mtrref=undefined</a>	0.68

	<a href="http://d&amp;gwh=7973905D6242465D0C400CC5DA93A330&amp;gwt=pay&amp;assetType=opinion">d&amp;gwh=7973905D6242465D0C400CC5DA93A330&amp;gwt=pay&amp;assetType=opinion</a>	
	<a href="http://bbc.com/news/world-us-canada-42574139">bbc.com/news/world-us-canada-42574139</a>	0.92
	<a href="http://foxnews.com/story/2007/09/21/heritage-foundation-leveling-with-pakistan-on-afghanistan.html">foxnews.com/story/2007/09/21/heritage-foundation-leveling-with-pakistan-on-afghanistan.html</a>	0.81
	<a href="http://breitbart.com/national-security/2017/03/28/pakistan-islamabad-constructing-fence-along-afghan-border/">breitbart.com/national-security/2017/03/28/pakistan-islamabad-constructing-fence-along-afghan-border/</a>	0.68
xinhuanet.com/english/2018-01/24/c_136919033.htm Bias : 0.75	<a href="http://newyorker.com/news/news-desk/a-significant-deal-between-the-u-s-and-israel">newyorker.com/news/news-desk/a-significant-deal-between-the-u-s-and-israel</a>	0.69
	<a href="http://nytimes.com/2018/01/03/us/politics/trump-tweets-nuclear-button.html?mtrref=undefined&amp;gwh=DD4472067A407A5828213D469AE408A7&amp;gwt=pay">nytimes.com/2018/01/03/us/politics/trump-tweets-nuclear-button.html?mtrref=undefined&amp;gwh=DD4472067A407A5828213D469AE408A7&amp;gwt=pay</a>	0.67
	<a href="http://bbc.com/news/world-us-canada-42553507">bbc.com/news/world-us-canada-42553507</a>	0.77
	<a href="http://foxnews.com/opinion/2011/09/16/time-is-now-for-three-state-solution-to-end-israeli-palestinian-conflict.html">foxnews.com/opinion/2011/09/16/time-is-now-for-three-state-solution-to-end-israeli-palestinian-conflict.html</a>	0.88
	<a href="http://breitbart.com/national-security/2017/03/30/leaders-of-egypt-jordan-and-palestinian-authority-to-meet-with-trump-in-april/">breitbart.com/national-security/2017/03/30/leaders-of-egypt-jordan-and-palestinian-authority-to-meet-with-trump-in-april/</a>	0.83
washingtonpost.com/politics/in-middle-east-trump-pence-shows-administration-is-unfazed-by-criticism-of-jerusalem-decision/2018/01/24/0eb99d30-0125-11e8-9d31-d72cf78dbee_story.html?utm_term=.9b57a537b857 Bias : 0.81	<a href="http://newyorker.com/news/news-desk/trump-sabotages-his-own-mideast-peace-process">newyorker.com/news/news-desk/trump-sabotages-his-own-mideast-peace-process</a>	0.81
	<a href="http://nytimes.com/2018/01/03/world/middleeast/trump-israel-palestinians-twitter.html?mtrref=undefined&amp;gwh=14D4AB4FB34FB4F9B44ABBC9389845DB&amp;gwt=pay">nytimes.com/2018/01/03/world/middleeast/trump-israel-palestinians-twitter.html?mtrref=undefined&amp;gwh=14D4AB4FB34FB4F9B44ABBC9389845DB&amp;gwt=pay</a>	0.72
	<a href="http://bbc.com/news/world-us-canada-42402350">bbc.com/news/world-us-canada-42402350</a>	0.82
	<a href="http://foxnews.com/politics/2018/01/02/trump-mulls-cutting-off-aid-to-palestinians-says-theyre-no-longer-willing-to-talk-peace.html">foxnews.com/politics/2018/01/02/trump-mulls-cutting-off-aid-to-palestinians-says-theyre-no-longer-willing-to-talk-peace.html</a>	0.58
	<a href="http://breitbart.com/jerusalem/2017/05/01/abbas-preps-trump-meeting-week/">breitbart.com/jerusalem/2017/05/01/abbas-preps-trump-meeting-week/</a>	0.85

After this, we compared the user generated rankings of relative bias of the given three articles, with those generated by our algorithm. The relative rankings for a dataset, both user generated and predicted by our method, are shown in table 7.

We chose the Kendall's Tau and Spearman's Rho rank correlation coefficients [19] for our comparison. The Spearman correlation score comes to be 0.54 and Kendall's Tau correlation as 0.47, indicating that the user rankings of bias correlated well with the bias scores generated by our algorithm.

TABLE VII. CORRELATION OF THE USER GENERATED RANKS WITH PREDICTED RANKS

Article URL	Aggregated User Rank	Predicted Rank
<a href="http://wsws.org/en/articles/2018/01/06/pak-i-j06.html">wsws.org/en/articles/2018/01/06/pak-i-j06.html</a>	3	5
<a href="http://newyorker.com/news/news-desk/the-tapi-pipeline-and-paths-to-peace-in-afghanistan">newyorker.com/news/news-desk/the-tapi-pipeline-and-paths-to-peace-in-afghanistan</a>	5	3
<a href="http://nytimes.com/2018/01/09/opinion/pakistan-trump-aid-engage.html?mtrref=undefined&amp;gwh=7973905D6242465D0C400CC5DA93A330&amp;gwt=pay&amp;assetType=opinion">nytimes.com/2018/01/09/opinion/pakistan-trump-aid-engage.html?mtrref=undefined&amp;gwh=7973905D6242465D0C400CC5DA93A330&amp;gwt=pay&amp;assetType=opinion</a>	1	1
<a href="http://bbc.com/news/world-us-canada-42574139">bbc.com/news/world-us-canada-42574139</a>	4	6
<a href="http://foxnews.com/story/2007/09/21/heritage-foundation-leveling-with-pakistan-on-afghanistan.html">foxnews.com/story/2007/09/21/heritage-foundation-leveling-with-pakistan-on-afghanistan.html</a>	6	4
<a href="http://breitbart.com/national-security/2017/03/28/pakistan-islamabad-constructing-fence-along-afghan-border/">breitbart.com/national-security/2017/03/28/pakistan-islamabad-constructing-fence-along-afghan-border/</a>	2	2

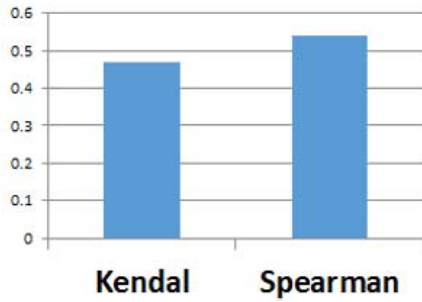


Fig. 6. Plot of the Kendall and Spearman rank correlation values for the aggregated user generated rankings and the algorithm generated rankings of bias score for the selected news articles

### C. Subjective reviews by reviewers

We also collected subjective reviews by the same ten users, on the overall usability of the news article recommendation system. Most of the users thought the system could be useful while browsing, though a couple of users thought they were not politically aware enough to judge how much biased the articles were.

Since the number of users is too small to be meaningful, in future we will repeat the study with more users.

## V. CONCLUSION AND FUTURE WORK

In this paper, we have described a system to recommend related URLs from the same topic from news sources with a variety of political biases. We hope that such a system can help to enable users to be more aware of bias in news articles they are currently reading. Our system offers a compromise

between keeping user autonomy and the need to warn the users of biased or fake news.

An early edition preprint of this paper has been uploaded to ArXiv [20] and the code is available on GitHub [21].

We intend to extend our work in a number of directions. First is to incorporate additional parameters such as similarity to other biased articles, context of the biased words and the number of superlative adjectives used.

Another is to use publicly available news trust or fake news identification APIs to reward or penalize the computed bias score.

Another is to examine the feasibility of searching relevant news articles quickly by using public search APIs.

Finally, sentiment analysis can be explored along with the bias score to develop a more composite score.

In future, we will also investigate state of the art techniques for news articles clustering such as contextual bandits.

## REFERENCES

- [1] Pew Research Center. Sharing the news in a polarized Congress. Dec 18, 2017. [Online]. Available: [people-press.org/2017/12/18/sharing-the-news-in-a-polarized-congress/](http://people-press.org/2017/12/18/sharing-the-news-in-a-polarized-congress/)
- [2] Erica Anderson, Google. Building trust online by partnering with the International Fact Checking Network. Oct 26, 2017 [Online]. Available: [blog.google/topics/journalism-news/building-trust-online-partnering-international-fact-checking-network/](http://blog.google/topics/journalism-news/building-trust-online-partnering-international-fact-checking-network/)
- [3] Margi Murphy, Telegraph. Facebook defends two-question fake news survey. 24 January 2018 [Online]. Available: [telegraph.co.uk/technology/2018/01/24/facebook-defends-two-question-fake-news-survey/](http://telegraph.co.uk/technology/2018/01/24/facebook-defends-two-question-fake-news-survey/)
- [4] Harper Neidig, the Hill. PragerU sues Google, YouTube for 'censoring' conservative videos. 24 Oct 2017. [Online]. Available: [thehill.com/policy/technology/356966-prageru-sues-google-youtube-for-censoring-conservative-videos](http://thehill.com/policy/technology/356966-prageru-sues-google-youtube-for-censoring-conservative-videos)
- [5] Anish Anil Patankar, Joy Bose. Bias Based Navigation for News Articles and Media. In Proc. NLDB 2016, pp 465-470.
- [6] Github. Bias\_senti, a Chrome extension that shows sentiment and bias of the current article. [Online]. Available: [github.com/patankaranish/bias\\_senti](https://github.com/patankaranish/bias_senti)
- [7] Pew Research Center. Political Polarization and media habits. October 21, 2014. [Online]. Available: [journalism.org/2014/10/21/political-polarization-media-habits/](http://journalism.org/2014/10/21/political-polarization-media-habits/)
- [8] Sailesh Kumar Sathish, Anish Anil Patankar, Nirmesh Neema. Semantics-Based Browsing Using Latent Topic Warped Indexes. InProc. Tenth International Conference on Semantic Computing (ICSC), 2016 IEEE
- [9] Bing Liu. Sentiment analysis and subjectivity. Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca. 2010.
- [10] Balahur A, Steinberger R, Kabadjov M, Zavarella V, Van Der Goot E, Halkia M, Pouliquen B, Belyaeva J. Sentiment analysis in the news. arXiv preprint arXiv:1309.6202. 2013
- [11] M. Recasens, C. Danescu-Niculescu-Mizil, D. Jurafsky. Linguistic Models for Analyzing and Detecting Biased Language. Proceedings of ACL, 2013.
- [12] Wikipedia. Mainstream media. [Online]. Available: [en.wikipedia.org/wiki/Mainstream\\_media](http://en.wikipedia.org/wiki/Mainstream_media)
- [13] D. M. Blei, A. Y. Ng, M. I. Jordan, "Latent Dirichlet allocation", The Journal of machine learning research, vol. 3, no. 2003, pp. 993-1022.
- [14] Brita Belli, Tsai Centre for Innovative Thinking at Yale. Yale Students Win Hackathon Challenge to Counter Fake News. December 12, 2017 [Online]. Available: [city.yale.edu/blog/2017/12/12/yale-students-win-hackathon-challenge-to-counter-fake-news](http://city.yale.edu/blog/2017/12/12/yale-students-win-hackathon-challenge-to-counter-fake-news)

- [15] Open Mind. Devpost. [Online]. Available:devpost.com/software/open-mind-rp60o2
- [16] B.S. Detector. A browser extension that alerts users to unreliable news sources. [Online]. Available:bsdetectortech/
- [17] Gensim: Deep learning with word2vec. [Online]. Available:radimrehurek.com/gensim/models/word2vec.html
- [18] T Mikolov, K Chen, G Corrado, J Dean. Efficient estimation of word representations in vector space. arXiv:1301.3781.
- [19] Wikipedia. Rank Correlation [Online]. Available:en.wikipedia.org/wiki/Rank\_correlation
- [20] Patankar AA, Bose J, Khanna H. A Bias Aware News Recommendation System. arXiv preprint arXiv:1803.03428. 2018
- [21] Github. Bias\_senti, a Chrome extension [Online]. Available:github.com/patankaranish/bias\_senti
- [22] Shuai Li, Alexandros Karatzoglou, and Claudio Gentile. Collaborative Filtering Bandits. ACM SIGIR 2016.
- [23] Gentile, Claudio, Shuai Li, Purushottam Kar, Alexandros Karatzoglou, Giovanni Zappella and Evans Etrue. On Context-Dependent Clustering of Bandits. ICML 2017.