International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

Analytical Look - Transparency in Online Informative Technology

Divyanshi Tyagi, Divyansh Khare, Disheak Ahlawat, Mansi, Gaurav Yadav

Abstract*: *We are predicting the bias (left leaning or right leaning) in online news articles based on text of online articles collected and the publication. The rating (either Left or Right) was assigned by looking at the leaning of the publication as found on mediabiasfactcheck.com. As we know the importance of online news has evolved with the advancement in technology. In order to understand the biasness in online journalism related to text of an article, we used a deep Neural Net to make classifications based on the labeling assigned according to publication. If the political bias of the publisher creeps in and such a correlation is there the AI will be able to learn it. Our training produced a very accurate classification model. This shows that online media is not as transparent when presenting news. The methodology is described ahead.

Keywords*: *Classification, Journalism, Neural Net, online news articles, Polarity.

I. INTRODUCTION

A. Background

Objectivity is most important principle in the profession of journalism and one whose standard has degraded in an Internet oriented and often deeply partisan media environment, from blogger to a major news article can disseminate information to public. Understanding the polarity of news articles influencing the perception of that media source by any given reader presents interesting implications today. Neural networks and algorithms filter the news presented to any reader often without the knowledge of reader looking for how to reach the objective of articles of news.

Revised Manuscript Received on March 30, 2020.
* Correspondence Author
Divyanshi Tyagi, Pursuing Bachelor of Technology in Computer Science and Engineering from Meerut Institute of Engineering and Technology, Meerut.
Divyansh Khare, Pursuing Bachelor of Technology in Computer Science and Engineering from Meerut Institute of Engineering and Technology, Meerut.
Disheak Ahlawat, Pursuing Bachelor of Technology in Computer Science and Engineering from Meerut Institute of Engineering and Technology, Meerut.
Mansi, Pursuing Bachelor of Technology in Computer Science and Engineering from Meerut Institute of Engineering and Technology, Meerut.
Gaurav Yadav, Pursuing Bachelor of Technology in Computer Science and Engineering from Meerut Institute of Engineering and Technology, Meerut.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Despite the importance of objectivity in journalism to integrity and perception of a given articles, previous researches of machine learning to automatically classify online news articles based on the presence of biasness in journalism in an article has been surprisingly sparse, especially detecting the biasness with in articles not specifically focused on political campaigns and figures[1]. Here in paper, we propose an approach for the strategy of classifying online news with the help of keyword level analysis which facilitates the polarity of the news by defining whether it is biased or unbiased and also whether it is left or right leaning based on the text.

B. GOAL

Using combination of survey data which assigns polarity to a publisher whether left or right and news articles collected from leading publishers of both sides ,
We are predicting:

I. The biasness in journalism of given online news articles based on text of the given article.

Our goal is to better understand, how biasness relates to the text of article as well as political learning influences the objectivity of the article.

II. DATA SETS AND FEATURE PROCESSING

As these is some existing research on this topic and no prior fixed data set so, we have created new data set with different articles from various online and offline sources and bias labels were assigned by using well know portals.

A. Article Text Collection

We have collected 320 online news articles on the basis of keywords and top ten Google News search [4]. Keywords are chosen to reflect a range of popularity of nonpolitical and political influences across categories. The complete list of keywords is provided in the Appendix.
Article texts are preprocessed using Python NLTK [5] to remove the punctuation and normalize texts of the news articles. We then filtered articles with less than 50 words and articles with videos are not considered to contain enough text for a reader to consistently judging the biasness of the text.

B. Label Assignment

The main task was to get the bias labels for the articles that we had collected .We had only collected articles from established publications like New York Times, Brier Bart, New York post and Atlantic.

The first thought that came to mind was to get readers to assign ratings but it was not possible. So the Labels were assigned according to the publisher. These publishers have been declared Left wing and Right wing according to their views. Our aim was to analyze if this biasness creeps into the articles. Articles from Left wing publications were assigned a bias rating of 1 and Right wing publications were assigned a rating of 0.

III. PREDICTION MODELS AND METHODOLOGY

A. Publication Recognition

We first took article's text only to predict the publisher of that article to find out if the semantic rules are clearly distinguishable among publishers .

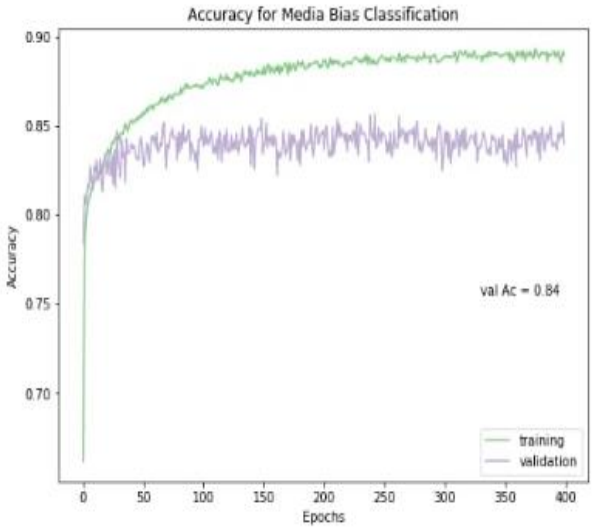
Classification of data is done by using a Deep Neural Network designed in Keras. The Neural Network had 4 softmax nodes in the final layer to make 4 way predictions on the 4 Publications.

Various techniques were considered for text vectorization. We tried the tf-idfvectorizer from SKLearnand got 3200 features which were deemed too high. In the end we settled on Google's Universal Sentence Encoder module from Tensor flow-Hub. This modules is a pre trained Network that victories text to [1x512] length vector.

All the articles were converted to 512 length vectors. The loss function used with the neural network is Sparse Categorical Cross Entropy and the optimizer is Adam .The network was trained for 400 epochs.

B. Bias Prediction

We have used earlier data set of same text. In this phase the goal is to predict the rating of an article's bias (left or right). We ran another neural net on this dataset. It was similar in architecture to the one used for publisher prediction but the last layer now had two nodes only, one for left and the other for right. The loss function and the optimizer were also kept same.



IV. RESULTS

Using the final set of the text features obtained after feature selection using the module from Tensor flow Hub; we have made predictions for the articles being biased towards the Left or Right.

1. We experimented with different Neural Network architectures to find the best performing one.
2. The results shown are for the best performing model.

V. CONCLUSION

We are able to predict the bias labeling, relatively the rough binary classification, with high accuracy, by the performance of the Neural Net .This indicates the presence of some biasness or lean in an article which is correlated to the political alignment of the publisher . Apart from the variation in semantic rules of each publisher, there exists some degree of biasness associated with an article. Overall, our research shows that it is absolutely possible to analyze and predict the polarity on the basis of bias labeling to article based on their article text. The results also provide insight on how thinking can influence the biasness of the news articles, blogs etc. being written on internet and disseminate information.

APPENDIX

Search Keywords: Selected articles are based on top Google news result searches for terms, based on Google trends: Trending 2018-19, Top Politicians, Trending Personalities, Political Agendas, Elections, Government, Top Leaders, Top International and National News, Laws, Terrorism, Defense, Economy. Search keywords: 'budget 2018', 'modi', 'congress', 'pulwama', 'brexit', 'rafale', '#me too', 'elections 2019', 'rahulgandhi', 'yogi', 'article 370', 'scst act', 'cow preservation', 'Christ church shootings', 'nitingadkari', 'kathua', 'budget 2019', 'india us tariff revise', 'nirmalasitharaman', 'us syria relation', 'immigration laws us', 'nick jonas', 'niravmodi', 'statue of unity', 'section 377', 'arvindkejriwal', 'priyaprakashvarrier', 'karnataka election results', 'nitishkumar', 'iranindia us'.

REFERENCES

1. Media Bias chart, https://www.adfontesmedia.com/
2. Budak, C., Goel, S., & Rao, J.M. Fair and balanced? Quantifying media bias through crowdsourced content analysis. (2016) Public Opinion Quarterly, 80(S1), 250-271.
3. Sonal Gupta. Finding bias in political news and blog websites.
4. http://snap.stanford.edu/class/cs224w-2010/proj2009/report SonalGupta.pdf, 2009.
5. Muhammad Usama Islam, Polarity detection of online news articles based on sentence structure
6. https://www.researchgate.net/profile/Muhammad_Usama_Islam/publication/323064270_Polarity_detection_of_online_news_articles_based_on_sentence_structure_and_dynamic_dictionary/links/5c6122c992851c48a9c9445e/Polarity-detection-of-online-news-articles-based-on-sentence-structure-and-dynamic-dictionary.pdf
7. Albert Chu, Prediction of Average and Perceived Polarity in Online Journalism
8. http://cs229.stanford.edu/proj2014/Albert%20Chu,%20Kensen%20Shi,%20Catherine%20Wong,%20Prediction%20of%20Average%20and%20Perceived%20Polarity%20in%20Online%20Journalism.pdf
9. Google Inc. Google trends 2018 https://trends.google.com/trends/topcharts?date=2018.
10. Steven Bird, Edward Loper and Ewan Klein. Natural Language Processing with Python. OReilly Media Inc.
11. https://beckernick.github.io/oversampling-modeling/

AUTHORS PROFILE

Divyanshi Tyagi is pursuing Bachelor of Technology in Computer Science and Engineering from Meerut Institute of Engineering and Technology, Meerut affiliated to Dr. APJ Abdul Kalam Technical University, Lucknow. She is a highly motivated person who likes to work in the field of computer science. She also works as a volunteer for social works and feels blessed to help the poor and unprivileged people. After graduation she wants to be a satisfied Software Engineer and travel adventurous places.

Mansi is pursuing Bachelor of Technology in Computer Science and Engineering from Meerut Institute of Engineering and Technology, Meerut affiliated to Dr. APJ Abdul Kalam Technical University, Lucknow. She is a highly motivated person who wants to serve the nation by joining Indian Armed forces. She also works as a volunteer for social works and feels blessed to help the poor and unprivileged people. After graduation she wants to be a part of Indian Army to serve the nation by all her deeds.

Divyansh Khare is pursuing Bachelor of Technology in Computer Science and Engineering from Meerut Institute of Engineering and Technology, Meerut affiliated to Dr. APJ Abdul Kalam Technical University, Lucknow. He is determined and likes to work in the field of machine learning and neural networks. His passion is to work with statistical data. After Graduation he wants to do masters keeping machine learning as the major.

Divyanshi Tyagi is pursuing Bachelor of Technology in Computer Science and Engineering from Meerut Institute of Engineering and Technology, Meerut affiliated to Dr. APJ Abdul Kalam Technical University, Lucknow. He is a highly motivated person who likes to work in the field of computer science. He also wants to work in as field where he can learn more about the artificial world. After graduation he wants to be a successful Engineer.

Disheak Ahlawat is pursuing Bachelor of Technology in Computer Science and Engineering from Meerut Institute of Engineering and Technology, Meerut affiliated from Dr. APJ Abdul Kalam Technical University, Lucknow. He is a highly motivated person who likes to work in the field of computer science. He also works as a volunteer for social works and feels blessed to help the poor and unprivileged people. After graduation he wants to be a successful Software Engineer and fulfill his dreams.