BIAS IN NEWS ARTICLE

Ankit Dahiya - MT19004, Gowtham G - MT19011, Jain Piyush Pradeep - MT19122

Motivation

The increase in media houses leads to an increase in new articles and social media makes it even easier for people to access these news articles. On the contrary, these news articles may not be true and may contain some bias towards a particular section of people. It is, therefore, necessary to find these biases to get the factuality of the news article being published.

Problem Statement

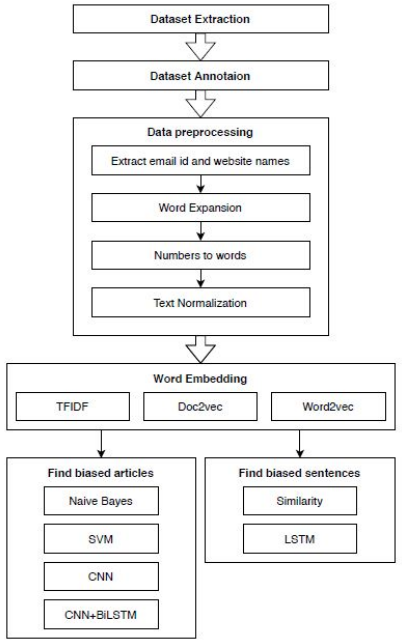
Finding the biases in the news article sentences towards the Indian National Congress or Bharatiya Janata Party, two of the major parties of India. Along with this find the sentences in which bias is present.

Dataset Used

New article dataset scraped using the news API and newspaper library in python. Articles are annotated with 1 (Support BJP), 0 (Neutral), -1 (Against BJP).

Information	Value
No of articles	2876
No of different sources	139
Average words in article	553
Articles Annotated	905

Methods



Preprocessing steps like numbers to word, word expansion are applied to the dataset.

Models based on Naive Bayes and SVM with the TF IDF vector were first used for classification.

Due to the low accuracy of these model CNN model with five convolution layer of different filter sizes (2,3,4,5,6) was build on where word2vec is used as embedding vector

CNN is good at extracting the features from the text. Bidirectional LSTM (BiLSTM) is used for maintaining the chronological ordering of data. So, CNN+BiLSTM model was the next model used with Doc2vec embedding vectors.

Two methods were used to find sentence bias namely cosine similarity and LSTM. LSTM performed better with an accuracy of 58%. The input vector is of size 150 which is converted using Doc2Vec and is passed to LSTM layer of size 512 with dropout 0.2 followed by a dense layer of size 3 with softmax function.

Results

Accuracy of various models for our dataset -

Model	Accuracy	F1_Score	Precision	Recall
NB	53.84	42.34	49.21	49.36
SVM	57.12	52.75	56.47	53.00
CNN	64.67	61.71	61.65	64.88
CNN+BiLSTM	84.44	83.02	86.27	81.81

CNN model performs better than NB and SVM models. Then in CNN+BiLSTM model, advantages of CNN and BiLSTM are both utilized which helped in achieving in good accuracy on test set.

LSTM result for sentence bias detection is shown below.

Model	Accuracy	F1_Score	Precision	Recall
LSTM	58	52	64	52

References

Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., & Nakov, P. (2018). Predicting factuality of reporting and bias of news media sources. arXiv preprint arXiv:1810.01765.