**Review of Fake-News Detection Techniques**

Neha Koppikar Sanika Mhadgut Salonii Tari

J027 J031 J051

1. **Introduction**

In the recent years, there has been a widespread of fake news at a phenomenal rate and this is not limited to just a single domain. This issue is everywhere: politics, healthcare, security, economics and so on. This has affected people financially, emotionally and has also led to people losing their lives. The major reason is that most of us consume news from world-level attention sources, which are the social media platforms. These sources have many advantages and disadvantages.

Examples of fake news include:

* Economics - Stock prices depend on a lot of factors. One of the factors include the company’s reputation. If someone wants to manipulate the stock prices, all that is required is spreading rumors about the reputation of a company and waiting for the results. This is probably what Harshad Mehta did to hike up the price of Associated Cement Company (ACC) from 200 INR to 9000 INR in 1990[2]. [1][2] Another example could be when 139 billion USD was wiped out when the Associated Press’ hacked Twitter account tweeted falsely about an explosion at the White House injuring President Barack Obama. This news was a rumor and therefore fake.
* Politics - [3] Former President Donald Trump has been accused of psychologically manipulating voters at the time of vote by showing potential voter’s good things about Donald Trump and malpractices done by Hillary Clinton on their Facebook feed. [3]
* Health - A lot of fake news got spread on WhatsApp regarding the corona virus that the disease did not affect youngsters and moving freely outdoors was fine. This news was fake because it failed to inform people that this was hazardous to non-youngsters in the house and transmission was possible (this information was released at a later point of time). The news was also proved incorrect when youngsters did in fact get infected with the virus.
* Security - It has been observed on all social media platforms that there are posts/messages giving the user incentive to perform a certain action resulting in accounts being hacked or our data privacy is breached. Since we lose something instead of gaining, this is yet another example of fake news.

The major disadvantage is the speed at which fake news gets communicated and oftentimes those are the ones which reach a wider audience as opposed to the real news. This is a huge challenge and a lot of industries are working to combat it. As a result, a lot of [1] fact-checking systems have been developed to fact-check textual claims. There has been an estimated increase of 400% in the number of fact-checking outlets in 60 countries since 2014. Snopes.com and politicalfact.com are examples of fact-checking systems. Among the companies who have invested in such systems are Facebook and Google. Facebook incorporated third-party fact-checking sites to social media posts and Google integrated fact-checking articles to their search engine. [1]

Fact-checking systems have definitely made things better. In this millennial and Gen Z world, people are generating a lot of data. Humans generate data that amount to trillions of bytes of data every day and a lot of this data is in text format (being the easiest to generate). Since the text is deliberately written to fool readers, it is difficult to detect fake news based on just the content. The content is directly sent to devices without any references.

Over the past decade, political campaigns have been increasingly waged on news sites and there have been ongoing concerns regarding how one’s political opinions and bias affect the information people read online. Though the current methods are fast, automated, and scalable compared to the cumbersome manual classification technique, there are only a few approaches to systematically analyze media bias. Bias classification has always been a widely debated and disputed topic which requires lots of context and current information about the political system which the existing machine learning models fail to identify and thus perform poorly on unseen data.

In cognitive science, bias is defined as deviation from the norm, or true value [1]. Media bias refers to a deviation or skewness that is introduced by journalists, reporters and writers while describing a story or an event. Articles written by newspapers often tend to reflect the author’s inherent point of view especially in the political spectrum. When the general public gets fed up with such biased news or information, supporting or opposing their political viewpoints, it creates a dogma that hinders them from subjective rational thinking and right decision making. Thus, it has become increasingly important to check for the bias in any form of literature.

Bias at an article level can be identified by the authors choice of word, headlines, syntax and semantics. Use of spin words and phrases, words giving a dramatic effect or words that portray someone in a negative light indicate bias. Examples of such words are serious, Refuse, Crucial, turn up the heat, Critical, Offend, even though, Monumental, Erupted etc. Successful identification of such words is one of the basic methods to detect bias. Extensive research has been done on topics like geographical over reporting, sentiment analysis, fake news detection, and variation in intensity of coverage. These topics are closely interlinked with bias detection.

1. **Literature survey**

Existing research in the NLP community largely concentrates on sentiment analysis, fake news detection, topic classification for news articles and media bias detection primarily with LDA, multinomial Naive Bayes, clustering algorithms, and support vector machines [4]. Newer models like LSTM and Convolutional Neural Networks (CNN) have also shown promising results in the field of natural language processing. One such approach used was to find the polarity of each individual word in the article and gauge its overall average sentiment. The extremely positive or negative articles were classified as biased. Recasens et. al. detected language bias [6]by using a word dictionary curated by Wikipedia editors to ensure articles conform to Neutral Point of View (NPOV) rules. A team of student researchers from Yale University developed a plugin called OpenMind to counter Fake News.[7] The plug-in used existing sentiment analysis technology to analyze an article. However, this project did not give information about the amount and type of bias present in the article. BS Detector, a Chrome extension with similar functionality used a curated list of unreliable news sources to flag online articles as being fake news or otherwise unreliable. Here the detection of bias was based on reliability of the source of the article rather than the content. Iyyer et al developed a Recursive Neural Network (RvNN) model to create an Ideological Book Corpus (IBC), which labels sentences and phrases based on their political ideologies. Their neural network was proven to outperform contemporary methods such as bag-of-words (BOW) models and hand-designed corpuses.

Building a supervised model for bias detection requires a data set with article-level labeling. The approach of manually labeling and classifying each news article as left, right and least biased requires a lot of resources and is oftentimes not feasible. For training the model itself the minimum requirement of labeled samples is in the thousands. Class imbalance also has to be taken into consideration while creating the data set. Bias identification may be highly subjective to the topic and person thus verifying articles will commonly require considerable time as compared to an expert. As the data set creation itself takes a lot of time most of the available labeled articles are outdated for analyzing contemporary articles, due to dynamic shifts in trending societal themes, discussion topics and news cycle. When the model is trained on outdated data it was observed that it performed well on the labeled test set however prediction on current news articles was not accurate. An alternative approach to creating labels was through distant supervision, where labels were generalized based on the source. Labeling approach derived from sites such as mediabiasfactcheck.com and allsides.com app has been shown promising in recent misinformation detection work by Horne et al [8]. The traditional Machine Learning and Neural Network approaches do not use reasoning about previous references in the article to decide if there is any bias at later points in the article. RNN’s addressed this issue by having networks with loops in them, allowing information to persist. LSTM are a special kind of RNN, capable of learning long-term dependencies.

1. **Existing Methods**

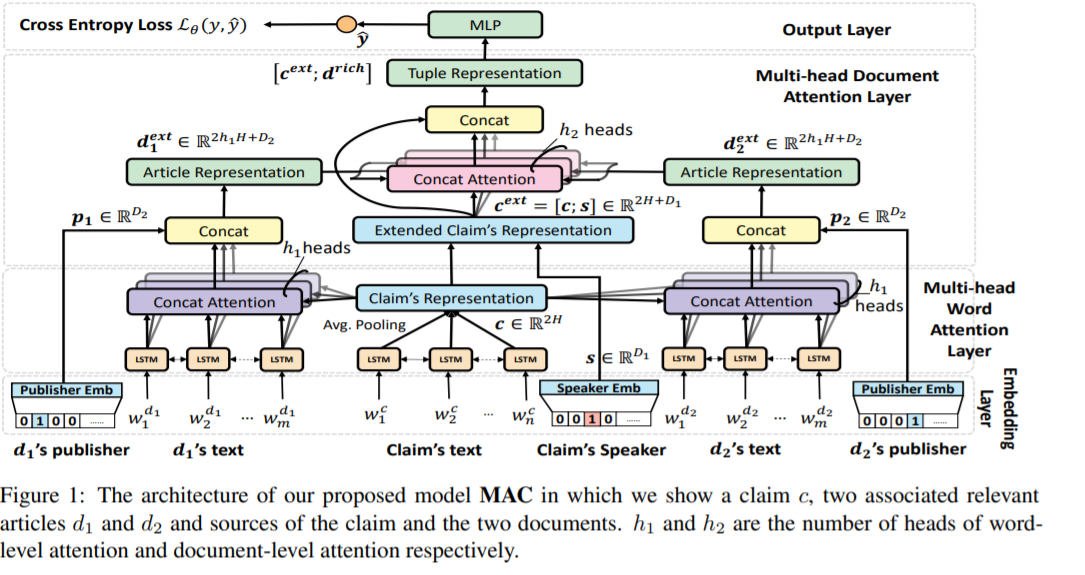
Fake news is not a recent problem. It has been affecting people and industries since a really long time. To solve this problem, researchers in 2011 proposed to use linguistics and textual content. This leads us to the next problem: On social media platforms, the text messages are shared without any references. This makes it difficult to trace back and check the credibility of the information. Therefore, researchers decided to check the credibility based on temporal spreading patterns (2018, 2020), network structures (2018), and user’s feedback (2019).

However, there are a few posts or messages that have links to external sources (mostly websites) to know more about the information sent. Therefore, recent work started developing in this area. In 2018, Popat et al. used word-level attention to all documents giving them equal importance. In 2019, Ma et al. only focused on crucial documents without considering what words help explain credibility of textual claims.

Observing drawbacks of the previous work, Vo et al. [1] proposed a Hierarchical Multi-head Attentive Network which jointly utilizes word attention and evidence attention. Every document will be written and read from a different point of view. That means there is no one way to interpret what has been written. Multi-head document-level attention mechanism helps in capturing contributions from different frames of references. This attention mechanism also takes speakers’ and publishers’ information into consideration. Contributions include:

* Proposing a novel hierarchical multi-head attention network that combines word-attention and evidence-attention which combines understanding from the words as well as the evidence to detect fake news.
* Proposing a novel multi-head attention mechanism that aims to capture important words and evidence considering different semantics.
* Testing on 2 public datasets and demonstrating their effectiveness as compared to other existing baseline methods.

1. **Methodology Explained**



The framework of this model has 4 main components:

1. Embedding Layer
2. Multi-head word attention layer
3. Multi-head document attention layer
4. Output Layer

**4.1 Embedding Layer**

Each piece of information is modelled as a sequence of words and the document is viewed as another set of words. All the words in the news will be converted into a d-dimensional vector and all the words in the document will be converted into an embedding matrix with the vocabulary size as the number of rows and d as the number of columns (d is the number of words in the news article). Each speaker and publisher are transformed into numerical categories using the one-hot encoding technique, which are further transformed into dense vectors using the matrices formed. These matrices are formed using the bag of words made using the publisher data and speaker data as corpuses. The model formed from the publisher data and the speaker data are not the same. The models are created separately from their respective data. Both the data for the models are uniformly initialized as an array. One thing to note: Both the models are jointly learned with other parameters of the Hierarchical Multi-Head Attentive Network for Fact-checking.

**4.2 Multi-head Word Attention Layer**

The word embeddings generated from the claim (news) are used as an input into a bidirectional LSTM which help generate the contextual representation for the hidden states in the forward and backward pass of BiLSTM. This approach is used for multiple documents so as to interpret the claim from a variety of semantics. This requires a series of operations which also includes normalization by the softmax operation.

**4.3 Multi-head Document Attention Layer**

The attention layer has three layers namely, the extending representations of claims, the extending representations of evidence, multi-head document attention mechanism. A multi-head attention mechanism is used to generate different attention distributions which constitute diverse contributions of articles. These series of operations require normalization by softmax operation. This method captures different semantic contributions of evidence integrated from speakers and publisher’s information towards the fact-checking process.

**4.4 Output Layer**

Tuple representation of the textual claim along with speaker, publisher and related documents information which is obtained from the document-level attention layer is fed into a neutral network. It will output the probability y-hat that the textual claim made by that speaker is not fake news. The neural network’s last layer has a sigmoid activation neuron to output probability distribution between true news and fake news. We increase the accuracy of the model by minimizing error i.e., distance between the networks estimation and ground truth thus triggering a feedback loop to change parameters of the MLP (mini batch gradient descent) The datasets used for testing this novel architecture are collections of tuples including textual claims and their corresponding ground truth labels collected from two fact-checking websites - snopes.com and politifact.com. Snopes did not have speaker’s information. The articles relevant to the textual claim c are obtained by scouring through the internet. Claims in Snopes dataset were labeled as true or false while in the PolitiFact dataset true, mostly true and half true got labelled together as True claims [9] and the rest into false claims.

1. **Results and Discussion**

Automated Fact-Checking systems often fare poorly due to the significant obstacle of bias identification- one which requires considerable context (about the societal norms and past knowledge) and its role in current atmosphere. Bias introduced by authors and hidden agendas can be detected through use of certain words having different sentiments, dramatic phrases giving a better representation of the contents thematic spread and their objectivity.

The MAC methodology proposed by Vo et al [1] when tested on data collected from two major fact-checking websites showed its fake news-detection performance exceeded existing baselines significantly. Using the one-sided paired Wilcoxon test on PolitiFact and Snope datasets with different parameters, the proposed methodology gave by far the best results. BERT which uses only textual content, HAN, DeClare both of which use both textual and relevant article’s text/content to fact-check textual content had low scores. Techniques which use document-level attention like HAN, LSTM-Avg, LSTM-Last lacked attention mechanism that gives ranked importance to words to better represent textual content and spread within a document. DeClare fared better than these methods by utilizing word-level attention. In both cases i.e., when positive class is true news and when positive class is fake news, there was an average improvement of 5 % over baselines by the MAC methodology. Two level attention mechanism fares better as better representation is built of documents/evidence because of the stacked layers like a pyramid building on the previous layer.

1. **Conclusion**

Media has a tremendous influence on thoughts, ideas and decisions of people thus it is essential that news media should be fair and accurate. Fast-checking systems have solved the problem of manually checking the credibility of news and doing this manually will take forever. The easier solution is to automate the process. This paper is an attempt to solve this problem using Artificial Intelligence and review existing methods to fact-check textual claims. Independently assessing the impact of attention layers to models performance yielded similar results as shown in performance comparison described above. Document-level attention independently fares poorly due to giving equal importance to irrelevant information leading to poorer representation of content as compared to word-level attention. Independently testing the MAC model on different combinations of data sources showed that combination of text and speakers and publishers sources gave best results. Improvements in MAC performance by decreasing heads in document-level and increasing heads in attention-level layer decisively proves that word attention plays a more important role in fact-checking than evidence attention. Future scope can include other sources like images/gifs/continuous threads and other information about the poster to improve performance.

1. **References**

[1] Hierarchical Multi-head Attentive Network for Evidence-aware Fake News Detection, author - Vo, Nguyen and Lee, Kymin, year - 2021

[2] Wikipedia: Harshad Mehta

[3] [What did Cambridge Analytics do during the 2016 Election?](https://www.npr.org/2018/03/20/595338116/what-did-cambridge-analytica-do-during-the-2016-election)

[4] Martha Bellows. Exploration of classifying sentence bias in news articles with machine learning models. 2018.

[5] Stephan Greene and Philip Resnik. More than words: Syntactic packaging and implicit sentiment. InProceedings of human language technologies:The 2009 annual conference of the north american chapter of the association for computational linguistics, pages 503–511. Association forComputational Linguistics, 2009.

[6] Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky.Linguistic models for analyzing and detecting biased language. InPro-ceedings of the 51st Annual Meeting of the Association for ComputationalLinguistics (Volume 1: Long Papers), pages 1650–1659, 2013.

[7] John Merriman Sholar and Noa Glaser. Predicting media bias in onlinenews cs 229: Machine learning-final project. 2016

[8] NELA-GT-2018: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles

[9] Kashyap Popat, Subhabrata Mukherjee, Andrew Yates, and Gerhard Weikum. 2018. Debunking fake news and false claims using evidence-aware deep learning. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 22–32.