AUTOMATIC IDENTIFICATION OF FAKE NEWS

(TERM PAPER)

SUPERVISOR: **Dr. Gagandeep Kaur**

SUBMITTED TO: **Dr. Sandeep Kumar Singh**
Dr. P. Raghu Vamsi

GROUP MEMBERS:

Nikita Jain	**14103010**
Saumya Pandey	**14103078**
Aditi Bhardwaj	**14103108**

Automatic Identification of Fake News

Saumya Pandey, Nikita Jain, Aditi Bhardwaj, Gagandeep Kaur

Department of CSE & IT
Jaypee Institute of Information Technology
Noida, India – 201307

Abstract The issue of "deliberately deceptive" or fake news has recently become a potential threat to high-quality journalism and well-informed public discourse. In the wake of this issue we were motivated to develop neural stance detection models based on machine learning and natural language inference-- i.e. identifying whether a particular news headline "agrees" with, "disagrees" with, "discusses," or is unrelated to a particular news article -- in order to allow people to more easily identify possible instances of "fake news". In this paper, we would first leverage an SVM trained on TF-IDF cosine similarity features to know whether a headline-article pairing is related or unrelated. If we classify the pairing as related, we would then employ various neural network architectures built on top of Long-Short-Term-Memory Models (LSTMs) to label the pairing as agree, disagree, or discuss. Ultimately, our goal is to find best performing neural network architecture to achieve high accuracy while identification of fake news.

1 Introduction

The spread of fake news articles has generated noticeable concern recently, as false or misleading stories can spread faster and reach a wider audience over social media. Fake news are online stories that appear to be factual but are not. They may appear in websites that appear to be legitimate, although often they are often found on websites with little or no real news. Unfortunately, sometimes fake news stories are picked up by legitimate media houses and appear on their websites. Given that rapidly changing technology has played a major role in enabling the spread of fake news, a natural question is whether technology can also help warn users about misleading news [1]. While technology is certainly not advanced enough to evaluate the truth of a claim on its own, it could be used to aid journalists and make it easier for them to detect and debunk false statements.

While automated fact checking and stance detection has not yet gathered much attention from researchers, previous research in Natural Language Inference (NLI) has worked on problems that are very similar to ours. NLI attempts to identify the relationship between two statements, by identifying whether two bodies of text support, contradict, or are neutral towards one another. Researchers in NLI have achieved reasonable success on this task using neural network models, and almost all of the best performing models on the benchmark Stanford SNLI corpus have incorporated neural networks. Matching LSTMs were successfully used for NLI to achieve a performance of 86.1% on the SNLI dataset. Tests a LSTM model, attention model, and a word-by-word attention model on the SNLI corpus. In their paper, word-by-word attention achieves the highest test accuracy of 83.5% on the SNLI dataset. A sequential LSTM-based model combined with a syntactic parsing model was used by to achieve 88.3% accuracy on the SNLI corpus [2].

With the advent of fake news being used to influence elections and negatively impacting the decision making of the readers, the identification of fabricated information has become an important task. Governments, newspapers and social media platforms are working hard on distinguishing credible news from fake news. The goal of the Fake News detection is to automate the process of identifying fake news

by using machine learning and natural language processing. This process can be broken down into several stages. A first helpful step towards the identification of fake news is to understand what other news sources are saying about the same topic. That is why we initially focus on stance detection. Specifically, the task is to estimate the stance of a news headline, relative to the contents of a news article which can but does not have to address the same topic. Thus, the relative stance of each headline-article pair has to be classified as either unrelated, discuss, agree or disagree. The discovery of a disagreeing headline-article pair does not necessarily correspond to the discovery of a fake article, but it is an automated first step which could make human reviewers aware of a discrepancy.

It is said that "False travels around the whole world like wild forest fire"**.** With many individuals relying on the internet for daily news, fake news continues to be circulated on search engines and social media, which leads to inaccurate stories being virally shared worldwide. All of these fake articles being circulated contains outrageous headlines which were meant to attract the greatest amount of engagement from users, as well as pretending to be legitimate to gain credibility — at least at first glance[3]. This led to exponential engagement to millions of users through mindless sharing on social media such as Facebook. Social media has surged exponentially in last decade. The growing internet and smartphone users have given birth to new threat of Fake News on the social media. People spreading and blindly believing these news can have disastrous effects.

Social media has become an integral part of modern society as all are very keen to increase their social footprints and make their presence felt in the virtual communities and networks. But suddenly the very primary cause of this social media like creation and sharing of information, news, trends, best practices and opinions are being sidelined as the menace of fake news is popping up. The most prominent reasons for the fake news are:
1. Anti-social elements of the society are purposely spreading such fake news due to some vested interest and gaining benefits.
2. Some users who are blind followers of important persons like celebrities or political figures often don't take pain to check the authenticity of news and blindly do like/comment/share.
3. The present form of IT Act is not equipped with proper provisions to check the spreading of fake news.
4. Also the growing smart phones penetration can be hold responsible as it is extending the free access to social media through internet connectivity it offers[4].

2 Problem Statement

The spread of fake news articles has generated noticeable concern recently, as false or misleading stories can spread faster and reach a wider audience over social media. Given that changing technology has played a major role in enabling the spread of fake news, a natural question is whether technology can also help warn users about false claims. While technology is certainly not advanced enough to evaluate the truth of a claim on its own, it could be used to aid journalists and make it easier for them to detect and debunk false statements.

There are several applications which have been designed for this purpose. Our application will explore the use Natural Language Processing techniques to determine whether a body of text agrees, disagrees, discusses, or is unrelated to another. This model could be applied as an automatic fact checker that could read an article, and then find other articles that either disagree or agree with its content. For this we will use methods from the field of Natural Language Inference to build a model that classifies the relationship between a news article headline and the body of a different news article.

With the advent of fake news being used to influence elections, the identification of false information has become an important task. Governments, newspapers and social media platforms are working hard on distinguishing credible news from fake news. And for this people developed the Fake News Challenge with the goal to automate the process of identifying fake news by using machine learning and natural language processing. This process can be broken down into several stages. A first helpful step towards the identification of fake news is to understand what other news sources are saying about the same topic. This is a problem which needs to be dealt with so we have also applied some of the concepts of deep learning to detect fake news and stance.

3 Literature Review

Towards Automatic Identification of Fake News: Headline-Article Stance Detection with LSTM Attention Models [9]

Sahil Chopra, John Merriman Sholar, Saachi Jain
Department of Computer Science
Stanford University

New Media, including online newspapers and social media, not only enable people around the world easily receive news and share information in real-time, but also lets fake news and rumors spread quickly under no verification and time[20]. Mistaking fake news for authentic reports can have costly consequences, as being misled or misinformed negatively impacts our decision-making and its consequent outcomes[8]. Fake, fabricated, falsified, disingenuous, or misleading news reports constitute instances of digital deception or deliberate misinformation [8]. "Digital deception", a term signifying deception in the context of information and communication technology, is defined here as an intentional control of information in a technologically mediated environment to create a false belief or false conclusion[8].While human fact checkers fail to verify to verify the credibility of the news due to the sheer volume of the information out there leveraging AI and machine learning techniques seems the only possible solution for automatic fake news detection in this era[20].Therefore we examined several papers regarding this issue to get greater insight into the problem.

Firstly, we researched the papers that approached the problem of fake news via stance detection In February 2017, the Fake News Challenge 1 (FNC-1) was launched by a non-profit organization with the goal of developing tools to help fact checkers tag fake news [9].The papers utilizes the dataset provided by FNC-1 to approach this problem. We developed our baseline models based on detailed analysis of the papers which involved the main task of classifying the headline-article pair into agree, disagree, discuss or unrelated. Thus, stance detection for fake news is split into a two classification problem in which the first baseline model was a Lexicalized Linear Classifier. This classifier for each article,headline pair we extracted features such as the cosine distance between TF-IDF vectors, max BLEU score, cross-grams, and Jaccard Distance[9]. Then SVM classifier with a radial basis function (RBF) kernel is applied based on the highest correlation value.The researcher then employed various neural network architectures built on top of Long-Short-Term-Memory Models(LSTMs) to label the pairing as agree, disagree, or discuss[9]. Therefore, second baseline model was a BOW MLP that utilized 300 Dimensional GloVe Embeddings to represent the headline and article in vector space[9].Softmax classification was implemented on the final hidden state of LSTM with concatenated inputs that produced least accurate results which lead to author resorting to better models after pre-processing the dataset that involved truncating the articles to reduce the timesteps.

To improve the accuracy of results the author moved on to a conditionally encoded LSTM model that involves two separate LSTMs, one for the headline and one for the article. The headline is fed through the first LSTM to extract the final hidden vector hn. This hidden state is used to initialize the LSTM of the article, thus ”conditioning” the article LSTM on the headline[9] and then adding different forms of attention to the LSTM. The paper constantly evaluates the accuracy of each model and based on the results implements models that yield highest accuracy while taking into account the constraints like reduction of timestamps etc. [9]. As the culminating model in this series of LSTM-based architectures, the author implemented the Conditionally Encoded LSTM with Bidirectional Global Attention using Bidirectional RNNs (thus reading the text both forward and backward)[9].This was the best performing model among all the models which yielded highest accuracy and successfully dealt with stance detection for fake news.

While the author has thoroughly scrutinized the issue of fake news and conducted intensive research before applying the algorithms for stance detection, author restricted his scope of work only on to a single dataset of FNC. Moreover, author has clearly overlooked the aspect of working on multi-lingual dataset and has limited his work to a single language. Finally, on one hand the author acknowledges the shortcomings and the constraints but doesn’t provide a proper plan or solutions to tackle these constraints that might lead to greater accuracy on the dataset in the future.

Towards News Verification: Deception Detection Methods for News Discourse[10]

Victoria L. Rubin, Niall J. Conroy, and Yimin Chen
Language and Information Technology Research Lab (LIT.RL)
Faculty of Information and Media Studies
University of Western Ontario, Ontario, CANADA

Secondly, we researched the paper that identified deception in textual data by identifying verbal predictors of deception with text processing and machine learning techniques[10].The paper intensively focuses on what is fake news, explores and identifies the sources of fake news like citizen journalism as Citizen journalists are not obliged to follow the guidelines of source-checking and fact checking cultivated in professional journalism [11] which leads to creation and dissemination of unverified news and looking to develop solution for detection of deception in news.In this paper Rhetorical Structure Theory (RST) and Vector Space Modeling (VSM) are the two theoretical components used in the analysis of deceptive and truthful news[10]. RST is used to analyze news discourse and VSM is used to interpret discourse features into an abstract mathematical space [10].This model aims to identify rhetorical relationships between the more and the less emphasized texts in the articles functional relations among different meaningful text units, and describes a hierarchical structure for each story [12].While the VSM each news text is represented as vectors in a high dimensional space [13], [14]. Then, each dimension of the vector space is equal to the number of rhetorical relations in a set of all news reports under consideration[10].The vectors represented in space are put into 2 clusters that is truthful and deceptive The element of a cluster is a news story, and a cluster is a set of elements that share enough similarity to be grouped together, as deceptive news or truthful news [15].The paper utilized the datasets of the National public radio show “Bluff The Listener” which was in the radio announcement style.The researcher performs in depth analysis by describing the procedures involved in RST analysis like-Inter-coder reliability test methods and calculations with consequent data manipulations, Statistical Procedures for Predictive Modeling using R etc [10].This was followed by application of logistic regression model for testing the accuracy on the training and the test set to calculate the accuracy achieved.

While the RST-SVM clustering technique for the NPR's "Bluff the Listener" news reports was only in part successful (63% accuracy), further steps need to be taken to find predictors of deception for a news verification system as predictive model is comparable to average human lie detection abilities (54% accuracy)[10].Thus the results were promising but inconclusive and better models and research will be needed to achieve higher accuracy in the future[10].Therefore, author needs to deploy neural network models and machine learning to move a step further in the news verification problem.

Stance Detection for the Fake News Challenge with Attention and Conditional Encoding [19]

Stephen Pfohl, Oskar Triebe, Ferdinand Legros
Department of Computer Science
Stanford University

Fake news poses a serious threat to our society nowadays, particularly due to its wide spread through social networks. While human fact checkers cannot handle such tremendous information online in real time, AI technology can be leveraged to automate fake news detection. The first step leading to a sophisticated fake news detection system is the stance detection between statement and body text.

By stance detection actually we refer to the task of classifying a pair of sources of text as agreeing, disagreeing, neutral, or unrelated. As a result, the focus of the challenge is to instead develop a stance detection tool that is able to automatically reason about the relationship between a new claim or headline and a set of articles. The stance detection task introduced by Fake News Challenge aims to identify perspectives of news body texts toward headlines.

In this research paper researchers have actually tried out to find a link between the headline and its corresponding body by applying various neural network algorithms such as RNN (Recurrent neural networks) and some of them have even used LSTM to identify fake news. This research paper firstly involved the pre-processing of the dataset that they had and then applying different models and seeing the expected outcomes. They analyzed their dataset using various models such as bag of words and LSTMs and attention algorithms and predicted the results. They have applied all the algorithms and every time any one of the algorithms prove to work better than the previous one. The attention models outperformed the simpler models on all performance metrics on the test set. In particular the models with neural attention were able to achieve significantly higher F1 scores predicting the infrequent stances Agrees and Disagrees. The main disadvantage found in this research paper was basically that the length of the article was different for different algorithms which means we need to pre-process the data every time before applying the algorithm which I think in my opinion is time consuming and requires a lot of effort .

Neural Stance Detectors for Fake News Challenge [20]

Qi Zeng, Quan Zhou, Shanshan Xu
Department of Physics
Stanford University

In this research paper they have basically applied different forms of Recurrent Neural Networks and analyzed their results. Recurrent neural networks basically have loops in them, allowing information to persist. A recurrent neural network can be thought of as multiple copies of the same network, each passing a message

to a successor. The main task focused upon on Stance detection is identifying good features to find out the relation between headline and body text. If we are given a pair of headline and body text, the goal is to identify whether the body text is related to the headline or not, and in which of the four categories the exact relationship between them lies. Four different labels were assigned to each headline-body pair: Unrelated, Discuss, Agree and Disagree. In this research paper a score metric is also calculated for the task of stance detection. Since relationship identification plays a major role in detection of fake news, a correct prediction or if the headline and the body text are related carries a major portion of the score, while discrimination over related/unrelated is less weighted. More specifically, 25% weighted score is for correct prediction on relatedness, and another 75% weighted score is for correct prediction on relationship.

A direct bidirectional encoder was able to achieve more accurate prediction results for agree category, though there was a lot of leakage of misclassification in discuss category. By using Attentive Reader with full attention, the recall on agree category was improved significantly, though the precision was still not high enough. In all cases, they notice that disagree category was poorly predicted. Their result suggests that all neural network based methods outperformed the hand-crafted feature based approach. In addition, they tried to improve the existing Attentive Reader with a full attention mechanism between words in body text and headlines. Evaluation result shows that this algorithm gave the best performance over other explored methods.

To further improve the performance, one needs to first improve the precision on agree category, and further improve the accuracy on disagree category. One possible approach is the combination of hand-crafted features and hidden features learnt by RNN to find more accurate results in the disagree category as this category provided the least accurate results. Their results in all of these categories were considerably good except the disagree category which somehow had a negative impact as the main task to be addressed was to find unrelated headline and the body text.

RNNs for Stance Detection between News Articles [2]

Jason Yu Chen, Joseph Johnson, Graham Yennie
Stanford University
Stanford, CA 94305

The spread of fake news articles has generated noticeable concern recently, as false or misleading stories can spread faster and reach a wider audience over social media. Given that changing technology has played a major role in enabling the spread of fake news, a natural question is whether technology can also help warn users about false claims. While technology is certainly not advanced enough to evaluate the truth of a claim on its own, it could be used to aid journalists and make it easier for them to detect and debunk false statements.

Through their research paper they have tried to explore how to use Natural Language Processing a techniques to determine whether a body of text agrees, disagrees, discusses, or is unrelated to another. The bag-of-vectors model worked very well on detecting related/unrelated news articles (93.9% accuracy), but not as well on detecting disagreement, agreement, and discussion. On the other hand, the RNN with attention model did not do as well at detecting relatedness, but did detect relationships relatively well. A cascading model that incorporates the Bag of Vectors model for relatedness, and then incorporates a attention model for the other categories could potentially improve our score. Perhaps an ensemble approach would improve the score as well [2].

The main disadvantage that we particularly found in this research paper was that the length of the article bodies had an upper quartile limit of roughly 460 words and the headlines had a limit of roughly 16 words. Although capturing more words in the body should equate to capturing more meaning in the model, later in their future work they will be discussing on how to apply a lower cut-off to the body length in the networked models improved the performance of the model and improve the score.

Stance Detection for Fake News Identification [5]

Damian Mrowca, Elias Wang, Atli Kosson
Department of Electrical Engineering
Stanford University
Stanford, CA 94305

With the advent of fake news being used to influence people, the identification of false information has become an important task. Governments, newspapers and social media platforms are working hard on distinguishing credible news from fake news. The goal of this paper is to automate the process of identifying fake news by using machine learning and natural language processing. This process can be broken down into several stages. A first helpful step towards the identification of fake news is to understand what other news sources are saying about the same topic and analysing the results.

The main task focused in this research paper is to identify relationship between headline and the body-text. To address this problem, they have begun with stance detection, which is considered to be a first step towards identifying fake news. The goal of this project is to identify whether given headline-article pairs: (1) agree, (2) disagree, (3) discuss the same topic, or (4) are not related at all. The method they have applied feeds the headline-article pairs into a bidirectional LSTM which first analyzes the article and then uses the acquired article representation to analyze the given headline.

To their surprise, results were quiet astonishing, processing the article first and conditioning the headline on the article encoding worked better than vice versa for their dataset. Just by switching the order in which the article and headline are processed, they were able to increase their performance and F1 score.In the future, they want to experiment with other recurrent architectures that have proven to be successful in natural language processing. Since the current network performs already almost perfectly on distinguishing unrelated from related articles, they believe that investigating methods that do particularly well at discriminating positive from negative stances would be most beneficial to improve their current model

References

[1] http://guides.library.usciences.edu/CurrentIssues/FakeNews

[2] *RNNs for Stance Detection between News Articles*
https://web.stanford.edu/class/cs224n/reports/2757443.pdf

[3] http://www.bbc.com/news/blogs-trending-37846860
[4] http://www.insightsonindia.com/2016/11/21/6-issue-fake-news

[5] *Stance Detection for Fake News Identification*
https://web.stanford.edu/class/cs224n/reports/2760496.pdF

[6] http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[7] https://talbaumel.github.io/attention/

[8]- J.T. Hancock, "Digital deception: When, where & how people lie online," *The Oxford handbook of Internet psychology*, A. N. Joinson, et al., eds., Oxford University Press, 2012, pp. 508.

[9]- *Towards Automatic Identification of Fake News:Headline-Article Stance Detection with LSTM Attention Models-https*://web.stanford.edu/class/cs224n/reports/2761028.pdf

[10]- *Towards News Verification: Deception Detection Methods for News Discourse* **Victoria L. Rubin, NiallJ. Conroy, and Yimin Chen-**
http://onlinelibrary.wiley.com/doi/10.1002/pra2.2015.145052010082/pdf
[11] D. Gillmore, "We the media: The rise of citizen journalists," *National Civic Review*, vol. 93, no. 3, 2004, pp. 58-63.
[12] W.C. Mann & S.A. Thompson, "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization,"*Text*, vol. 8, no. 3, 1988, pp. 243-281.

[13] C.D. Manning & H. Schutze, *Foundations of Statistical Natural Language Processing*, The MIT Press, 1999.

[14] G. Salton & M.J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, 1983.

[15] P. Berkhin, "Survey of clustering data mining techniques.," 2002; DOI: 10.1.1.18.3739.

[16]-UCI Machine Learning | Kaggle -https://www.kaggle.com/uciml
|https://www.kaggle.com/mrisdal/fake-news/data

[17]- http://research.signalmedia.co/newsir16/signal-dataset.html

[18]- *Stance Detection for the Fake News Challenge:Identifying Textual Relationships with Deep Neural Nets-https*://web.stanford.edu/class/cs224n/reports/2760230.pdf

[19]- *Stance Detection for the Fake News Challenge with Attention and Conditional Encoding-https*://web.stanford.edu/class/cs224n/reports/2748568.pdf

[20]-*Neural Stance Detectors for Fake News Challenge-*
https://web.stanford.edu/class/cs224n/reports/2761936.pdf,