Detecting fake news using hybrid convolutional neural networks

Eliza Vardanyan
Earlham College
Richmond, Indiana
evarda17@earlham.edu

ACM Reference Format:
Eliza Vardanyan. 2020. Detecting fake news using hybrid convolutional neural networks. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Social media has become the platform to share ideas, arts as well as news. The opportunity to engage a wider audience has attracted news channels and freelance reporters with almost no cost and fast. However, more significant issues arise from this, such as the source's unreliability, the spread of misinformation among social media users, and public opinion manipulation. The urge of controlling misinformation has become extremely necessary with the ongoing pandemic. Misinformation about various treatment methods, 'facts' about COVID-19 made it difficult for the healthcare workers to fight against the virus[4].These are the reasons why detecting fake news on social media is of great importance and relevance. Fake news is defined by Cambridge dictionary as "false stories that appear to be news, spread on the internet or using other media, usually created to influence political views or as a joke." Systematic literature review shows that the following are the most common approaches to detecting fake news: Language approach, Topic Agnostic approach, Machine Learning approach, Knowledge Based approach, and Hybrid approach[2]. Language approach is based on the linguistic analysis of the text and has three main sub-approaches: Bag of Words, Semantic Analysis and Deep Syntax. Bag of Words is based on identifying frequency of each word in the textual context as a sign related to misinformation, while Semantic Approach focuses on the sentiment features of the text. In contrast, Deep Syntax analyzes the syntax structure of the text and compares that to a pre-established, known pattern of lies and uses this as a sign to inform about the truthfulness or fakeness of the text. Topic-agnostic approach takes into consideration the topic-agnostic features such as number of ads on the page, morphological patterns in the text, percentage of total semantic words in text, the difficulty of understanding the text, and the pattern of the layout of the web page's[1]. Machine learning approaches use datasets to train algorithms and detect fake news. Knowledge based approach uses external sources of news for the verification purposes of the text. Hybrid approach uses a combination of " human and machine learning to help identify fake news" [2]. This takes into account three elements of news articles: text of the news/article, users' response to the article and the source of the article. I will be exploring a hybrid approach of detecting fake news but with the possibility of integrating new features outside of the three main elements that are already established in the field of hybrid approach of fake news detection. In this paper, I will be looking at natural language processing (NLP), deep geometric learning and convolutional neural networks (CNN) for textual analysis and feature extraction. For image reliability, I will be reviewing MediaEval Verifying Multimedia. Note that image reliability refers to the detection of fauxtography. This, according to Urban Dictionary, is also known as Fraudulent photography. Urban dictionary defines it as "News images that have been faked by various means, generally to promote an ideological agenda or to manipulate the emotions of the viewer."

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 DATASETS

This section looks at the datasets that researchers used in their works of fake news detection. In this section, I will explore the reason for certain choices for datasets and programs that were used as a benchmark for news comparison and identification as unreliable. In this section I will also discuss some specifics regarding the labeling that was used in the datasets, the filters of elimination as well as the size of the datasets.

2.1 Monti et al. Twitter Dataset

The dataset used in this paper consists of 1084 labeled claims that were spread on Twitter "in 158951 cascades covering the period from May 2013 - January 2018."[3] The number of unique users involved in this spreading was 202,375. The Following features were used when describing the news, users, users' activity. As a result, the dataset was grouped into four categories: user profile, user activity, network and spreading, and content.

2.2 Ruchansky Twitter and Weibo Datasets

This paper uses "two real world social media datasets that have been used in previous work, Twitter and Weibo." The following statistics were available on the datasets retrieved from Twitter and Weibo: number of users that were taken into consideration for short news statements on Twitter were 233,719 while this number on Weibo is 2,819,338. The number of articles of the datasets that were reviewed in this paper was 992 on Twitter and 4,664 on Weibo. The ratio of fake and true articles on both platforms were 1:1 [5].

Table 1: Textual Sources and Tools

Source	Authors
Twitter	Ruchansky et al.[5] Wang et al.[8] Zhang et al.[11] Monti et al.[3] Wang
Weibo	Wang et al.[8]
PolitiFact.com	Ruchansky et al.[5] Wang et al.[8] Zhang et al.[11]
Other (FakeNewsNet,Nopes, Buzzfeed)	Shu et al.[6] Monti et al.[3]

2.3 FakewNewsNet

Shu et al. uses FakeNewsNet that is resulted from collective fact-checking platforms called BuzzFeed and PolitiFact. The labels that are used in categorizing the publisher-news relation are five: “left”, “left-center”, “least-biased”, “right-Center”, and “right”. Only the publishers with the annotations of [“left”, “least-biased”, “right”] were chosen for the research for accuracy reasons. These labels were rewritten as [-1,0,1] to construct vectors [6].

2.4 LIAR

Wang uses here is called LIAR, which resulted in obtaining 221 statements from CHANNEL 4 and POLITIFACT.COM. LIAR dataset consists of 12.8K human labeled short statements from POLITIFACT.COM's API. The dataset was labeled under the following six labels: pants-fire, false, barely-true, half-true, mostly-true, and true (write each of these in “” quotes) [7].

2.5 Zhang Twitter Dataset

The datasets used in Zhang et al. consists of 12055 tweets fact-checked by PolitiFact. Six credibility labels were used but their associated 6 numerical scores were used instead like the following: “True”:6, “Mostly True”:5, “Half True:”4, “Mostly False”: 3, “False”:2, “Pants on fire!”: 1 [10] .

2.6 McIntire's Fake News Dataset

Zhou et al. “Generates adversarial examples from articles in McIntire's fake-real-news-dataset”([11]) that is very commonly used in the field of misinformation research. The dataset consists of 6335 articles. 3171 articles from this dataset were labeled as real and 3164 were labeled as fake.

2.7 LIAR and Yang Twitter Dataset

This paper,as a dataset, uses LIAR and uses BuzzFeed News to evaluate the algorithm performance. BuzzFeed contains up to 1627 news articles while LIAR consists of over 12000 short news statements. The paper also used “Twitter's advanced API with the tiles of news to collect related news tweets."[9]. After elimination of duplicates and filtering out non-verified users' tweets, the resulting news for LIAR were 332 and for BuzzFeed were 144 news. User response was used for the sentiment analysis and signage for the fake news detection as well. Other filtering techniques included eliminating tweets that had less than 3 engagement records (likes, retweets, etc.) since this way tweets that were not credible were eliminated.

3 METHODS

This section describes the methods that were used by researchers in the field of fake news detection. These methods were used to extract features from text and media as well as methods that were used to make the prediction.

3.1 Text Analysis

Wang et al.[7] represent LIAR: "a new, publicly available dataset for fake news detection from the surface-level linguistic pattern analysis". The analysis is done by using hybrid convolutional neural networks. This approach has its limitations because the neural networks introduced in this paper can only improve the effectiveness of those news statement that are solely text. LIAR is one of the most extensive datasets for fake news detection. It contains 12836 short news statements that were manually labeled. This dataset has the following filters applied to it: logistic regression, support vector machine, long short-term memory networks, and convolutional neural network model. Each statement in the dataset was fact-checked by the PolitiFact.com website. For the second stage of verification, a random subset of analysis reports were taken and checked weather there was an agreement between the analysis and the news. This dataset was created to assist in the development of an automatic fake news detection model. The paper uses the CNN model because of the previous work done in the field and the exceptional results. The metadata was later encoded with an initialized matrix-vector. The author then compared the CNN model with other well-known methods for fake news detection such as SVM (Support Vector Machine), and bi-directional long short-term model (Bi-LSTM). CNN performed the best compared to these methods. Overfitting was an issue in the Bi-LSTM method, while SVM and Logistic Regression (LR) models with the proposed dataset had performance improvements.

Zhang et al.[10] introduced FAKEDETECTOR, which focuses on deep diffusive network models. This approach is unique with its four component analyses of the dataset. First, it completes an article credibility analysis that includes textual content, then it does creator credibility analysis, goes over creator-article publishing historical records, and computes subject credibility analysis. The paper reviews the word cloud of 'true' and 'false' statements in the first component analysis. It uses words that differentiate true and false news/statements as signals for future distinguishing purposes. These are part of the Explicit Feature Extraction steps in this paper. The next step in this analysis that again distinguishes this work from related work in the field is the Latent Feature Extraction. This step goes further, exploring hidden signals related to the publishers, statements, creators, authors. This is done with the use of deep recurrent neural network models. The result of the RNN model is then used as an input for Deep Diffusive Unit and Gated Diffusive Unit models. These models draw correlations between creators, authors, publishers, articles, and the content's subject.

Zhou et al. [11] is unique in its approach to the problem with its observance of fact-tampering attack method of detecting fake news. The paper highlights the importance of linguistic characteristics analysis in fake news detection. The proposed idea mainly focuses on linguistic features without doing fact-checking of statements. To visually portray this approach, the authors propose a knowledge graph known as the Straw Man Solution Approach. This idea of knowledge graphing via Crowdsourcing is widespread among well-known companies such as Google and Reddit. The paper is significant in fake news detection since it reviews all the most common and effective methods in the field and highlights the benefits and the defects of these methods. This paper shows that, in the real-world, linguistic characteristics, if observed alone, are not as effective. However, when combined with other fake news detection filters they can significantly improve accuracy.

Monti et al.[3] introduce a new approach to fake news detection based on deep geometric learning. This method relies on convolutional neural networks and graphs. "The model consists of four-layered Graph CNN that are two-dimensional and two fully connected layers" responsible for fake news prediction and classification. This paper is distinguishable from other related work because it discusses URL-based and cascade-based fake news detection approaches. Cascade is the term used for "the news diffusion tree produced by a source tweet referencing a URL and all of its retweets" in Monti et al. [3].

4 CONCLUSION

The literature review discussed how different authors approached text analysis for fake news detection. First, I explained what type of datasets have been used by researchers and the underlying reasoning behind the choice of certain kinds of data. Most of the researches used the Twitter dataset, while some also discussed Weibo. Almost all researchers used a non-profit independent journalist fact-checking website called PolitiFact.come for verification reasons. I also reviewed the methods that researchers used for text analysis, and CNN was the most common model used. I explained that this was the most common method reviewed by researchers because of its high prediction accuracy. The different approach and application of CNN was the distinguishing part of all the related works done in fake news detection. For future research, combining different techniques into one will be an interesting branch to look into to achieve higher accuracy.

REFERENCES

[1] Sonia Castelo, Thais Almeida, Anas Elghafari, Aécio Santos, Kien Pham, Eduardo Nakamura, and Juliana Freire. 2019. A topic-agnostic approach for identifying fake news pages. In *Companion Proceedings of The 2019 World Wide Web Conference*. 975–980.

[2] Matthee M. de Beer, D. 2020. Approaches to Identify Fake News: A Systematic Literature Review. Retrieved September 6, 2020 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7250114/

[3] Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M Bronstein. 2019. Fake news detection on social media using geometric deep learning. *arXiv preprint arXiv:1902.06673* (2019).

[4] Christina Pazzanese. 2020. Social Media Used to Spread, Create COVID-19 Falsehoods. Retrieved September 6, 2020 from https://news.harvard.edu/gazette/story/2020/05/social-media-used-to-spread-create-covid-19-falsehoods/

[5] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 797–806.

[6] Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 312–320.

[7] William Yang Wang. 2017. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648* (2017).

[8] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*. 849–857.

[9] Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu. 2019. Unsupervised fake news detection on social media: A generative approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 5644–5651.

[10] Jiawei Zhang, Bowen Dong, and S Yu Philip. 2020. Fakedetector: Effective fake news detection with deep diffusive neural network. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 1826–1829.

[11] Zhixuan Zhou, Huankang Guan, Meghana Moorthy Bhat, and Justin Hsu. 2019. Fake news detection via NLP is vulnerable to adversarial attacks. *arXiv preprint arXiv:1901.09657* (2019).