Deep Learning Approaches for Detecting Fake News

Subjects: Computer Science, Artificial Intelligence

Contributor: Pummy Dhiman , Amandeep Kaur , Celestine Iwendi , Senthil Kumar Mohan

The unregulated proliferation of counterfeit news creation and dissemination poses a constant threat to democracy. Fake news articles have the power to persuade individuals, leaving them perplexed. State of the art in deep learning techniques for fake news detection are described herein.

fake news detection deep learning

1. Fake News Detection Using Artificial Intelligence

A bibliometric study on 479 academic articles on fake news that were indexed in the Web of Science database was carried out using VOSviewer software ^[1]. There are three topologies provided: misinformation, mal-information, and disinformation ^[2]. Despite having similar sounds, these terms differ in some ways. Misinformation ^[2] is the unintentional dissemination of false information. It is not intended to deceive others. As an example, during COVID-19 times, the internet was inundated with information regarding precautions and home remedies. In contrast, disinformation ^[5] involves spreading bogus information with the intention of deceiving others. Contrarily, "misinformation" ^[2] is information that is grounded in reality but is utilized to hurt a victim, group, organization, or nation, e.g., revenge porn. It is recommended that further research be conducted on misinformation.

In order to focus on the use of an interdisciplinary approach contributing to scholarship and to comprehend the role of mobile and social media, as well as how the outcome of the 2016 US Presidential Election pushed the research community toward fake news detection, the authors in ^[4] used Google Scholar to examine 142 academically published articles. Google Scholar automatically keeps track of new articles as they are posted online. It has been found that prospective examiners should pay more attention to how people create, consume, and spread false information. To ascertain how the term "fake news" came to be a search subject, another bibliometric study explored 640 scientific articles that were taken from the WoS database ^[8]. These chosen articles were analyzed using Excel, SciMAT, and VOSviewer for scientific mapping. The authors came to the conclusion that educational practices must be promoted to stop the spread of false information.

Fake news goes by many names. The authors in ^[9] evaluated 387 fake news-related papers published on WOS between 2000 and 2019 to determine the overlap of and differences among phrases. They also recorded the distribution of publication years, author nations, and publications associated with these studies. This fake news study had nothing to do with people's views, fact-checking, or research on disinformation. Some studies on misinformation ^[10] simply discussed misinformation in their research fields.

Because traditional and online media were inundated with information about COVID-19, the WHO dubbed it an "infodemic". In 2020, one in six COVID-19 information pieces produced in India were false, making India the country from which the majority of COVID-19 false information originated ^{[5][11]}. Consequently, in January 2021, a concept map was created using 414 entries in the Web of Science (WoS) database, utilizing issues relevant to infodemics, misinformation, and disinformation to map the related literature ^[12]. Future research is also proposed with regard to useful research questions.

The rise of online media has hastened the spread of rumors and propaganda. Because these are forms of fake news, researchers have been motivated to work in this field. Some studies conducted bibliometric reviews to get a more in-depth look at this topic. The authors in ^[7] conducted an empirical study of peer-reviewed articles published in the WoS database from 1965 to 2019 and concluded that, despite an increase in publications and citations, propaganda cannot be considered an autonomous research field. Similarly, the authors in ^[11] used CiteSpace software to conduct a bibliometric review of 970 articles from the WoS database from 1989 to 2019 and concluded that rumors are spreading more quickly as a result of the development of various social networking platforms.

2. Deep Learning in Fake News Detection

Recent years have seen an explosion of research on fake news detection, which has resulted in numerous studies in the field. In the past, research on fake information detection focused on two main approaches: (a) categorizing false news in the context of unimodality (textual or visual); (b) exploiting multimodality (textual and visual) ^[13]. An overview of the use of DL in fake news detection is shown in **Figure 1**.



Figure 1. DL in news classification.

As it is known, computers cannot understand human language, and users express themselves in human language; thus, a subfield of AI known as NLP has emerged to fill this void. It enables machines to understand human communication languages. As a result, before using ML or DL, data should be preprocessed with NLP. The dataset is split into training and testing after this preprocessing step (2). A deep learning model is modeled using the

training set (step 4), while it is tested using the test set (step 5). Finally, in step 6, news is classified as either true or false.

Researchers from all over the world have helped to tackle the problem of fake news. An AI model using NLP and conventional ML was created to categorize classified text as fake news ^[9]. Experiments were carried out using the Kaggle dataset and articles pulled from the internet. When compared to other ML and DL models, the findings showed that random forest and CNN with global max pooling had high accuracy. Machine learning models were found to be roughly 6% less accurate than neural network models.

To indicate that people are more likely to forward news items immediately if the content is novel and emotionally appealing ^[13], the authors in ^[14] used a multilayer perceptron (MLP) neural network and a novelty and emotionbased representation of news to countermeasure the spread of fake news. During COVID-19 times, the internet was flooded with information related to home remedies and precautions; then, due to care, love, and fear of losing their loved ones, people shared this information through social media without checking whether it was true or false. This is how the spread of misinformation happens ^[15]. To implement this approach, four datasets were used, and statistically significant improvements in accuracy were recorded ^[14]. A hybrid DL framework (BerConvoNet) proposed by ^[16], combining BERT and CNN based on the news embedding block (NEB) and multiscale feature block (MSFB), was used to classify whether a given news article is genuine or not. However, it is yet to be explored how multimodal learning works.

A deep triple network (DTN) along with an entity-based attention network (EAN) developed in ^[17] can not only detect if a news article is true or false but also explain why it is. In this direction, the authors in ^[18] also attempted to work on a BERT-based model using two explanation methods, LIME ^[19] and Anchors, with the goal of detecting fake news while also attempting to explain it. The results supported the use of multiple surrogate-type methodologies to derive explanations using the Kaggle dataset, containing 44,848 news items. Because Anchors were not always able to provide an explanation, other explainability methods ^[7]9] will be applied in the future.

Another approach to solving the problem of falsified news was investigated in ^[20], i.e., a triple-branch BERT network for both binary and multilabel classification of fake news. Two datasets, LIAR and LIAR PLUS, were deployed; as compared with previously explored models, the accuracy was improved significantly using binary classifiers, but only marginally using multilabel classifiers.

GANS ^{[14][15]} are popular for content synthesis, i.e., for creating deep fakes. They consist of two architectures, a generator and a discriminator, competing against each other in multiple iterations to create more realistic content and to distinguish it from real content. In this context, a comprehensive review ^[21] was performed to analyze and identify existing approaches to deep fake creation and detection. Various deep fake tools such as Faceswap, DFaker, FSGAN, StyleGAN, and Face2Face, in addition to existing solutions based on the ML and DL techniques CNN, LSTM, RCN, and VGG16, as well as ResNet models and existing databases, were discussed.

To differentiate between fake (spoof) and real (bona fide) speech, a solution based on classical supervised learning pipeline was proposed ^[22]. The publicly available ASVspoof 2019 dataset was used to validate the proposed method. The authors of ^[23] used the FakeAVCeleb dataset to demonstrate that the ensemble-based method outperformed multimodal methods for detecting deep fakes of multiple modalities, i.e., video and audio. Along with deep fake movies, this collection also contained fake audio that was created artificially.

Examining the liaison between text and visual information made use of the cosine similarity notion. The processing of picture data was performed using a trained image2sentence model, and the representation of text representations was performed using a text-CNN with an additional fully connected layer ^[24]. F1 scores of 89.6% and 89.5% were recorded for the PolitiFact and GossipCop datasets, respectively. The authors of ^[25] used the latent Dirichlet allocation (LDA) to model online posts for fake news detection in an Indian context. Images were analyzed using VGG16 and Resnet-50, while text was analyzed using several ML and DL models. The implementation showed 74% accuracy.

To detect fake news, the authors created SpotFake ^[26]: a multimodal framework that combines language models with a VGG-19 model pretrained on ImageNet to incorporate contextual information. Using the concatenation technique, text and visual features were combined to form a multimodal fusion module. The investigation's findings presented 77.77% and 89.23% accuracy on publicly available datasets Twitter and Weibo, respectively. Inspired by ^[26], the authors proposed SpotFake+ ^[27], an advanced version of SpotFake, where transfer learning was used to extract semantic and contextual information from lengthy news articles and images.

References

- Zardari, S.; Alam, S.; Al Salem, H.; Al Reshan, M.; Shaikh, A.; Malik, A.K.; Ur Rehman, M.M.; Mouratidis, H. A Comprehensive Bibliometric Assessment on Software Testing (2016–2021). Electronics 2022, 11, 1984.
- 2. Ireton, C.; Posetti, J. Journalism, 'Fake News' & Disinformation; United Nations Educational, Scientific and Cultural Organization: Paris, France, 2018.
- Sehgal, V.; Peshin, A.; Afroz, S.; Farid, H. Mutual Hyperlinking Among Misinformation Peddlers. arXiv 2021, arXiv:2104.11694. Available online: http://arxiv.org/abs/2104.11694 (accessed on 30 March 2022).
- López-Robles, J.R.; Cobo, M.J.; Gutiérrez-Salcedo, M.; Martínez-Sánchez, M.A.; Gamboa-Rosales, N.K.; Herrera-Viedma, E. 30th Anniversary of Applied Intelligence: A combination of bibliometrics and thematic analysis using SciMAT. Appl. Intell. 2021, 51, 6547–6568.
- 5. Islam, M.R.; Liu, S.; Wang, X.; Xu, G. Deep learning for misinformation detection on online social networks: A survey and new perspectives. Soc. Netw. Anal. Min. 2020, 10, 1–20.

- Ilie, V.I.; Truica, C.O.; Apostol, E.S.; Paschke, A. Context-Aware Misinformation Detection: A Benchmark of Deep Learning Architectures Using Word Embeddings. IEEE Access 2021, 9, 162122–162146.
- Nightingale, S.; Farid, H. Examining the Global Spread of COVID-19 Misinformation. 2020, pp. 1– 12. Available online: http://arxiv.org/abs/2006.08830 (accessed on 13 July 2022).
- Van Eck, N.J.; Waltman, L. VOSviewer Manual; Universiteit Leiden: Leiden, The Netherlands, 2013; Available online: http://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.1.pdf (accessed on 15 June 2022).
- 9. Lai, C.M.; Chen, M.H.; Kristiani, E.; Verma, V.K.; Yang, C.T. Fake News Classification Based on Content Level Features. Appl. Sci. 2022, 12, 1116.
- Lee, S.; Nah, K. A Counterattack of Misinformation: How the Information Influence to Human Being. In Intelligent Human Systems Integration; AISC, no. Ihsi; Springer: Cham, Switzerland, 2020; Volume 1131, pp. 600–604.
- 11. Al-Zaman, M.S. COVID-19-Related Social Media Fake News in India. J. Media 2021, 2, 100–114.
- 12. WHO. Fighting Misinformation in the Time of COVID-19, One Click at a Time; World Health Organisation (WHO): Geneva, Switzerland, 2021. Available online: https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time (accessed on 11 January 2022).
- 13. Kaur, G.; Kaur, P.A.; Khurana, M. A Review of Opinion Mining Techniques. ECS Trans. 2022, 107, 10125.
- 14. Kumari, R.; Ashok, N.; Ghosal, T.; Ekbal, A. What the fake? Probing misinformation detection standing on the shoulder of novelty and emotion. Inf. Process. Manag. 2022, 59, 102740.
- 15. Rangra, K.; Kapoor, M. Exploring the mel scale features using supervised learning classifiers for emotion classification. Int. J. Appl. Pattern Recognit. 2021, 6, 232.
- 16. Choudhary, M.; Chouhan, S.S.; Pilli, E.S.; Vipparthi, S.K. BerConvoNet: A deep learning framework for fake news classification. Appl. Soft Comput. 2021, 110, 107614.
- 17. Liu, J.; Wang, C.; Li, C.; Li, N.; Deng, J.; Pan, J.Z. DTN: Deep triple network for topic specific fake news detection. J. Web Semant. 2021, 70, 100646.
- 18. Szczepański, M.; Pawlicki, M.; Kozik, R.; Choraś, M. New explainability method for BERT-based model in fake news detection. Sci. Rep. 2021, 11, 1–13.
- 19. Dieber, J.; Kirrane, S. Why model why? Assessing the strengths and limitations of LIME. no. iii. 2020. Available online: http://arxiv.org/abs/2012.00093 (accessed on 2 May 2022).

- 20. Mehta, D.; Dwivedi, A.; Patra, A.; Kumar, M.A. A transformer-based architecture for fake news classification. Soc. Netw. Anal. Min. 2021, 11, 39.
- Nguyen, T.; Nguyen, Q.H.; Nguyen, D.; Nguyen, D.; Huynh-The, T.; Nahavandi, S.; Nguyen, T.; Pham, Q.-V.; Nguyeng, C. Deep Learning for Deepfakes Creation and Detection: A Survey. SSRN Electron. J. 2022, 223, 103525.
- 22. Borrelli, C.; Bestagini, P.; Antonacci, F.; Sarti, A.; Tubaro, S. Synthetic speech detection through short-term and long-term prediction traces. Eurasip J. Inf. Secur. 2021, 2021, 2.
- 23. Khalid, H.; Kim, M.; Tariq, S.; Woo, S.S. Evaluation of an Audio-Video Multimodal Deepfake Dataset using Unimodal and Multimodal Detectors. In Proceedings of the ADGD 2021— Proceedings of the 1st Workshop on Synthetic Multimedia—Audiovisual Deepfake Generation and Detection, New York, NY, USA, 24 October 2021; co-located with ACM MM 2021. pp. 7–15.
- 24. Zhou, X.; Wu, J.; Zafarani, R. SAFE: Similarity-Aware Multi-Modal Fake News Detection; Springer International Publishing: Berlin/Heidelberg, Germany, 2020.
- 25. Sharma, D.K.; Garg, S. IFND: A benchmark dataset for fake news detection. Complex Intell. Syst. 2021, 7, 2199–2210.
- 26. Singhal, S.; Shah, R.R.; Chakraborty, T.; Kumaraguru, P.; Satoh, S. SpotFake: A multi-modal framework for fake news detection. In Proceedings of the 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), Singapore, 11–13 September 2019; pp. 39–47.
- 27. Singhal, S.; Kabra, A.; Sharma, M.; Shah, R.R.; Chakraborty, T.; Kumaraguru, P. SpotFake+: A Multimodal Framework for Fake News Detection via Transfer Learning (Student Abstract). In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 13915–13916.

Retrieved from https://encyclopedia.pub/entry/history/show/94895