

Social Media False Information Detection Based on AI and Machine Learning

Anindita Dey

Senior Program Manager, Strategy and Analytics at Meta
anindita09dey@gmail.com

Abstract: *The ubiquity of social media platforms presents both unprecedented opportunities and challenges. The world is evolving rapidly, and what is true one minute may or may not be accurate the next minute. There is vast amount of information out there and people may believe something is true when it is not. It is easy to create and modify text, images and even videos that may be used to spread misinformation. It may prove fatal in some cases. It is pivotal to stop the spread and remove misinformation that can potentially be consumed and spread across the globe. Therefore, there is a need to develop some mechanism to combat misinformation and fake news on the social media platforms. AI and ML plays a pivotal role in detecting misinformation or fake news as they can analyze data, identify trends and patterns of misinformation.*

Keywords: Misinformation, fake news, detection, prevention, social media, deep learning

1. Introduction

Misinformation is false information spread by users commonly known as bad actors via text, images or videos. Misinformation is spread and shared on social media during popular events that have a huge impact on our society such as elections, major sports event etc. The bad actors spread misinformation to deceive public, generate revenue through click baits or influence opinion of people on a certain topic. Recently, misinformation has been linked to major world events such as Covid 19 pandemic, 2016 US elections, Russia and Ukraine war, Israel and Palestine war etc. During these events, misinformation was shared widely leading to mistrust in public and had major consequences.

2. Materials and Methods

2.1 Machine Learning (ML)

Machine Learning is an AI technique that emphasizes on using data and algorithms so AI can behave the way humans' program, thereby increasing the efficiency. ML algorithms are designed to analyze data, their patterns and trends thereby allowing complex systems to make predictions or decisions based on the data. ML algorithms use statistical techniques and algorithms to train models on labeled or on unlabeled data, making the system to make accurate predictions on the data.

Machine learning is further classified into Supervised and Unsupervised learning. Supervised learning is a method where models are trained using labeled data with known output labels. In Unsupervised learning, the data is unlabeled, and the goal is to discover hidden patterns, structures or relationships.

2.2 Deep Learning Methods

Deep Learning uses neural networks to train computers. In Deep Learning, a hierarchy of layers is used to transform input data into a more abstract and composite representation.

This technique is commonly used to detect misinformation by analyzing the linguistic features of the content on social media.

2.3 Natural Language Processing (NLP)

NLP focuses on understanding and analyzing human language. NLP can classify text, extract data, analyze sentiments etc. NLP can analyze the language used in a post to identify inconsistencies, conflicts and biased language.

2.4 Network Analysis

AI can detect misinformation on social media by analyzing the network of users who are sharing the content on the platform. Network analysis can analyze patterns of engagements between users to identify fake news and monitor the spread of misinformation on social media. This helps social media companies in identification and detection of misinformation and thereby prevent it from spreading.

3. Literature Survey

3.1 Characteristics of Literature Survey

This paper aims to compile the latest advancements in using machine learning methods for identifying fake news. To achieve this goal, searches were conducted using Google Scholar, employing specific keywords such as "fake news detection," "social media," "data mining," "deep learning," "natural language processing." Papers meeting the criteria of employing machine learning techniques to address fake news issues and being written in English were included in the selection process.

3.2 Materials

In the below table, we can see summary of all the selected articles. The articles were classified based on data type and machine learning algorithms.

Table 1: Summary of Analyzed Papers

Articles	Data Type	Machine Learning Algorithms	Year
Shu et al., 2017 [1]	Categorization	Social media data mining	2017
Thota et al., 2018 [2]	Content-based data	Deep learning	2018
Monti et al., 2019 [3]	Network data	Deep learning	2019
Hirlekar and Kumar, 2020 [4]	Content-based data	Natural language processing	2020
Oshikawa et al., 2018 [5]	Content-based data	Natural language processing	2018

The below table provides a summary of the machine learning techniques discussed in each paper, categorizing them based on established machine learning technique classifications. Within each category, various techniques applied in the

papers are detailed. The detection approach encapsulates the essence of each paper, serving as a concise explanation from a machine learning standpoint, addressing the question: "How does this paper approach fake news detection?"

Table 2: Summary of Machine Learning Techniques used in the articles.

Articles	Main Line of Work	Techniques Used	Detection Approach
Shu et al., 2017 [1]	Social media data mining	Preprocessing and classification algorithms	Social media data mining
Thota et al., 2018 [2]	Deep learning	Deep Neural Networks	Neural network
Monti et al., 2019 [3]	Deep learning	Feature extraction	Geometric deep learning
Hirlekar and Kumar, 2020 [4]	Natural language processing	Sentiment Analysis	Natural language processing
Oshikawa et al., 2018 [5]	Natural language processing	Different Natural Language Processing techniques	Natural language processing

4. Results

Concerning social media data mining, this technique aims to utilize the extensive data accessible through social media platforms. It enables analysis of user behavior and network dynamics, offering valuable insights into explainability. This method elucidates anticipated user behaviors and the evolution of networks and users. In the realm of deep learning, which involves understanding intricate patterns within data, these algorithms hold immense promise. This proves advantageous in this scenario as they can grasp semantic and contextual details embedded within the provided data. Natural Language Processing enables the

detection of diverse linguistic patterns underlying fake news. It facilitates sentiment analysis, topic modeling, and textual examination, aiding in discerning whether a piece of news is authentic or not. However, solely relying on this technique may result in the loss of contextual information, as it primarily concentrates on text. Consequently, valuable data such as user-topic relationships and multimedia content might be overlooked. Furthermore, textual ambiguity poses a significant challenge; deciphering slang, irony, and sarcasm can prove difficult for these models.

The following table shows the merits and demerits of various techniques.

Table 3: Summary of merits and demerits of each technique.

Articles	Main Line of Work	Merits	Demerits
Shu et al., 2017 [1]	Social media data mining	- Large data volume - User behavior analysis	- Data quality and availability - Computation
Monti et al., 2019 and Thota et al., 2018 [2]	Deep learning Methods	- Learning complex patterns - Semantic and contextual information - Relational based features	- Labeled data - Computation
Hirlekar and Kumar, 2020 and Oshikawa et al., 2018 [4,5]	Natural Language Processing	-Leveraging linguistic patterns Enabling sentiment analysis, topic modeling, and textual analysis techniques	-Focuses only textual context. -Language ambiguity and context understanding
Monti, F.; Frasca, F [3]	Network Analysis	- Cost Optimization	Relies on accurate and reliable data

5. Conclusion

The articles examined on the detection of fake news have offered valuable perspectives on the progress and obstacles within this domain. Through the exploration of diverse methodologies such as data mining, deep learning, and natural language processing, researchers have achieved notable advancements in recognizing and addressing fake news circulating on social media platforms.

In summary, the discourse illuminated the advantages and disadvantages of different approaches to detecting fake news. While social media data mining provides valuable insights into user behavior and network dynamics, it grapples with issues like data quality, availability, and computational demands. Deep learning holds promise in comprehending intricate patterns and semantic details, yet it encounters

limitations due to the scarcity of labeled data, computational costs, and interpretability challenges. On the other hand, natural language processing excels in identifying linguistic patterns but may sacrifice contextual understanding and face difficulties with textual ambiguity.

In general, the realm of fake news detection is undergoing constant development, as researchers delve into diverse methodologies and pursue avenues to enhance performance, efficiency, and interpretability. The amalgamation of various approaches and optimization tactics is poised to foster the development of more potent and scalable solutions in the times ahead.

References

- [1] Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake News Detection on Social Media. *SIGKDD Explor.* **2017**, *19*, 22–36.
- [2] Thota, A.; Tilak, P.; Ahluwalia, S.; Lohia, N. Fake News Detection: A Deep Learning Approach. *SMU Data Sci. Rev.* **2018**, *1*, 10.
- [3] Monti, F.; Frasca, F.; Eynard, D.; Bronstein, M. Fake News Detection on Social Media Using Network Analysis.
- [4] Hirlekar, V.; Kumar, A. Natural Language Processing based Online Fake News Detection Challenges—A Detailed Review. In Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 10–12 June 2020.
- [5] Oshikawa, R.; Qian, J.; Wang, W.Y. A Survey on Natural Language Processing for Fake News Detection. In *Language Resources and Evaluation*; Springer Science+Business Media: Berlin/Heidelberg, Germany, 2018; pp. 6086–6093.