Investigations on Deep Learning Techniques for Detecting Fake News on Online Social Networks

Ramakrishnan Raman[1], Dr Syed Salman[2], Dr. Babasaheb Jadhav[3]

[1] Symbiosis International (Deemed University), Pune, India; raman06@yahoo.com

[2] Lincoln University College, Malaysia; syedahmed@lincoln.edu.my

[3] Dr. D. Y. Patil Vidyapeeth, Pune, India; babasaheb.jadhav@dpu.edu.in

---

Abstract: The widespread dissemination of fake news on online social networks poses a serious threat to public opinion, democratic processes, and societal trust. Traditional rule-based systems have shown limited success in combating this complex and evolving challenge. Deep learning (DL), with its ability to automatically extract high-level features from large-scale data, has emerged as a powerful tool in detecting fake news. This paper explores the application of deep learning models in fake news detection, focusing on techniques such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) networks, and Transformers. We examine various datasets, architectures, evaluation metrics, and the performance of different DL models in real-world scenarios. The paper also discusses challenges and future directions in building robust and generalizable fake news detection systems.

Keywords: Fake News Detection; Deep Learning; Social Networks; Accuracy, Rumors detection, classification

---

1. Introduction

Fake news refers to intentionally false or misleading information presented as news, typically to influence public opinion or obscure the truth. With the rise of online social networks like Twitter, Facebook, and Reddit, fake news can reach millions of users within minutes. Detecting fake news is challenging due to the speed of dissemination, the evolving nature of disinformation, and the subtleties in distinguishing fake from real news. Traditional machine learning techniques require manual feature engineering, which may not capture the complex semantics of textual data. Deep learning offers an end-to-end solution capable of learning intricate patterns directly from raw data [1] [2]. General model for fake news detection is shown in figure 1:

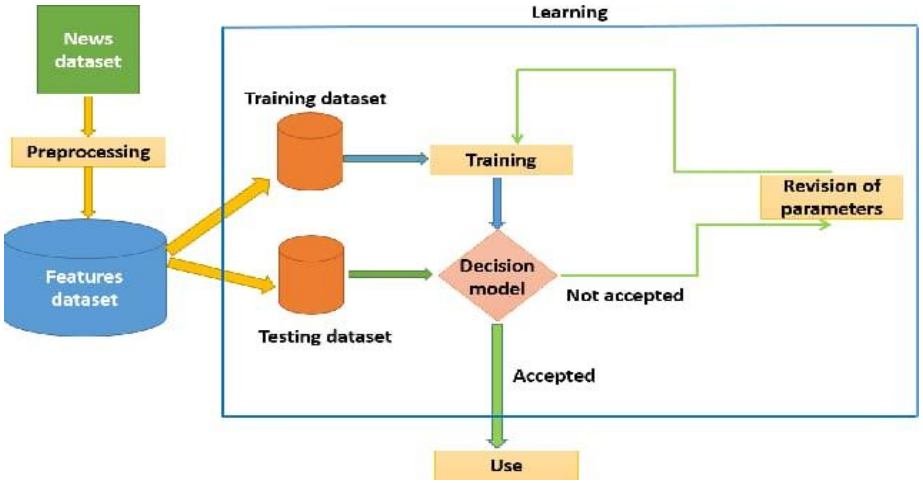

Figure 1: General model for fake news detection

Early approaches to fake news detection relied on metadata analysis, linguistic cues, and crowd-sourced fact-checking. With the advent of deep learning, research shifted toward using neural networks to learn representations of textual content[3]. CNNs have been employed to capture local patterns in text, while RNNs and LSTMs have been used to model sequential dependencies. More recent efforts leverage attention mechanisms and Transformer-based models such as BERT to capture contextual relationships. This section reviews significant contributions and summarizes their findings[4].

The detection of uncontrolled fake news by automated systems is a delicate skill that involves an awareness of social or political opinions in addition to "common sense." In spite of the fact that they appear as credible news providers, "fake news" websites are becoming increasingly difficult to identify. These websites are propagating content that is either intentionally misleading or completely incorrect. It is possible that fake news might bring about a decline in society as a result of the false information that it disseminates [4]. There is some validity to the assertion that various news items on television have an effect on the feelings of the general public. False information that is broadcast on news programs has the effect of diminishing the emotional experience of the audience as well as the sentiments of the general public. A feeling of betrayal will be experienced by many.
For the entire time, people are under the impression that they are acting and thinking in a certain way. If this problem is not adequately handled, it will interfere with the growth of society in a constructive and peaceful manner, and it will also lead to an increase in social instability [5]. Second, there is the possibility that television stations would suffer financial losses as a result of violations brought about by fake news. The credibility criterion is broken by false news, which means that it has the potential to swiftly grow into unlawful events. As a result, the program group and the television station might be subject to unpleasant lawsuit.
There is the potential for editors of news sources that publish fraudulent reports to be subject to requests for victim compensation and, in the most severe situations, criminal prosecution [6]. That is something that no one on the squad wants to be a witness to. The elimination of fake news to the greatest extent feasible is therefore of the utmost importance. On the other hand, the credibility of a political party or government is severely damaged when they publish fake news. The reputation of a television station suffers when they publish phone news. It is important that the data accurately reflect the perspectives and policies of both the government and the party. Various government entities are dependent on news broadcasts in order to accomplish their political and ideological objectives. Because of this, the news acts as a channel through which the nation's collective voice may be

conveyed. In the event that this issue is not addressed, the dissemination of misleading news poses a danger to both the legitimacy of government institutions and the good reputation of television networks.

Early approaches to fake news detection relied on metadata analysis, linguistic cues, and crowd-sourced fact-checking. With the advent of deep learning, research shifted toward using neural networks to learn representations of textual content. CNNs have been employed to capture local patterns in text, while RNNs and LSTMs have been used to model sequential dependencies. More recent efforts leverage attention mechanisms and Transformer-based models such as BERT to capture contextual relationships.

2. Related Work

Wang introduced the LIAR dataset, a large-scale benchmark composed of short political claims labeled by professional fact-checkers. This work laid the foundation for evaluating fake news detection models using real-world annotated data [5]. Yang et al. proposed TI-CNN, which combines text and image features using convolutional neural networks. This approach integrates content and context for better performance in multimodal settings [6].

Devlin et al. presented BERT, a powerful Transformer-based language model pre-trained on large corpora[7]. BERT's ability to learn bidirectional representations of text significantly boosted fake news detection tasks through fine-tuning. Volkova et al. focused on linguistic features and credibility signals in social media posts, applying stylistic and affective analysis to distinguish suspicious from trusted news on Twitter [8].

Ruchansky et al. introduced CSI, a hybrid model that combines user behavior, text content, and temporal features to detect fake news. CSI uses an RNN framework to model dynamic sequences in news propagation [9]. Kumar et al. explored disinformation on Wikipedia and highlighted the importance of structural features and editor behavior in detecting hoaxes [10].

Shu et al. developed FakeNewsNet, a comprehensive repository combining news content, user engagement, and publisher information. It enables the study of fake news within a social context [11] .

Vaswani et al. revolutionized NLP with the Transformer architecture, relying entirely on self-attention mechanisms. This architecture serves as the foundation for models like BERT [12] .

Zhou et al. introduced SAFE, a similarity-aware fake news detection method that utilizes multimodal and relational information to enhance accuracy [13]. Khattar et al. proposed MVAE, a multimodal variational autoencoder that fuses textual and visual cues to detect fake news more robustly [14]. Baly et al. contrasted news content with audience reactions, profiling media sources using both textual and contextual information [15].

3. Methods, data sets and Evaluation

3.1 Methods

- **Convolutional Neural Networks (CNNs)**

CNNs are adept at extracting local n-gram features and have been adapted for text classification by treating text as a sequence of word embeddings. In fake news detection, CNNs can identify specific phrases or patterns indicative of deception.

- **Recurrent Neural Networks (RNNs) and LSTM**

RNNs are designed to handle sequential data but suffer from vanishing gradient problems. LSTM networks address this issue with gating mechanisms, making them suitable for capturing long-term dependencies in textual data.

- **Transformer-Based Models**

Transformers, especially BERT and its variants, have redefined NLP tasks. Their ability to model bidirectional context using self-attention mechanisms allows for a deeper understanding of text semantics. Fine-tuning pre-trained Transformers has achieved state-of-the-art results in fake news detection.

- **Hybrid and Ensemble Models**

Combining CNNs, RNNs, and Transformers can leverage the strengths of each model. Ensemble learning methods such as stacking and boosting further improve robustness and generalization.

3.2 Data sets

Effective training of DL models requires large and diverse datasets. Commonly used datasets include:

- **LIAR**: Contains short political statements labeled by professional fact-checkers.
- **FakeNewsNet**: Integrates news content with social context information.
- **BuzzFeed News and PolitiFact**: Includes articles manually labeled as real or fake.
- **COVID-19 Fake News Dataset**: Focuses on health-related misinformation.

These datasets vary in content type, label granularity, and source credibility, impacting model performance.

3.3 Model Evaluation

Common metrics for evaluating fake news detection models include:

- Accuracy
- Precision, Recall, F1-score
- Area Under the ROC Curve (AUC)
- Confusion Matrix Analysis

Cross-validation and stratified sampling ensure reliable performance estimation. The choice of metric depends on the application context, such as minimizing false positives in public health scenarios.

4. Challenges and Limitations

Despite promising results, several challenges remain:

- **Data Quality and Bias**: Incomplete or biased datasets can lead to skewed models.
- **Generalizability**: Models trained on one dataset may not perform well on another due to domain-specific language.
- **Adversarial Attacks**: Manipulated inputs can mislead DL models.
- **Interpretability**: Understanding model decisions is crucial for trust and accountability.

5. Conclusion

Deep learning offers a robust framework for detecting fake news on online social networks. Through powerful representation learning, models such as CNNs, LSTMs, and Transformers can uncover subtle patterns in deceptive content. While current systems show high accuracy, addressing challenges related to generalization, transparency, and adaptability will be key to deploying trustworthy and scalable solutions. As social media continues to shape public discourse, deep learning will remain a vital tool in the fight against digital misinformation.

References

1. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.
2. Zhang, X., Ghosh, S., Dekhil, M., Hsu, M., & Liu, B. (2018). Automatic online fake news detection combining content and social signals. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2827-2836.
3. Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *Security and Privacy*, 1(1), e9.
4. Zhou, X., & Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. *arXiv preprint arXiv:1812.00315*.
5. Wang, W. Y. (2017). "Liar, liar pants on fire": A new benchmark dataset for fake news detection. *ACL*, 422-426.
6. Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2019). TI-CNN: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*.
7. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT*, 4171-4186.
8. Volkova, S., Shaffer, K., Jang, J. Y., & Hodas, N. (2017). Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter. *ACL*, 647-653.
9. Ruchansky, N., Seo, S., & Liu, Y. (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.

10. Kumar, S., West, R., & Leskovec, J. (2016). Disinformation on the web: Impact, characteristics, and detection of Wikipedia hoaxes. In *Proceedings of the 25th International Conference on World Wide Web*, 591-602.
11. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and dynamic information for studying fake news on social media. *Big Data*, 8(3), 171-188.
12. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *NeurIPS*, 5998-6008.
13. Zhou, X., Jain, A., Phoha, V. V., & Zafarani, R. (2020). Safe: Similarity-aware multi-modal fake news detection. In *Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM)*, 1398-1403.
14. Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). MVAE: Multimodal variational autoencoder for fake news detection. In *The World Wide Web Conference*, 2915-2921.
15. Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., & Nakov, P. (2020). What was written vs. who read it: News media profiling using text analysis and social media context. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2713-2724.