Context-Aware Fake News Detection Using Deep Learning Models

Jawaher Alghamdi, Yuqing Lin, and Suhuai Luo

School of Information and Physical Sciences, The University of Newcastle, Australia

The rapid proliferation of misinformation on social media has become a pressing issue, posing serious societal and political risks. Traditional fact-checking methods, which rely on manual verification and crowdsourcing, face significant scalability challenges, particularly in the era of big data. Furthermore, conventional automatic fake news detection (FND) methods, which primarily analyze textual content, often fail to capture contextualized information, limiting their effectiveness against sophisticated misinformation campaigns. To address these limitations, recent research has leveraged Pre-trained Language Models (PLMs) and Large Language Models (LLMs), which excel in capturing nuanced contextual information. Typically, these models are used in text classification tasks by adding an upper-layer classification head. In our prior study [1], we explored various PLM-based architectures and introduced novel approaches for FND. Our experimental evaluation on a real-world fake news dataset achieved SOTA performance, with an F1 score of 98%. Additionally, we conducted a comparative study [2] to assess the effectiveness of different ML models and PLMs across multiple datasets for FND. However, the high computational cost of fully fine-tuning these PLMs and LLMs presents a significant barrier to real-world deployment. Therefore, to mitigate the computational constraints of full fine-tuning, we proposed a parameter-efficient adapter-based approach (currently under review). This method involves freezing the PLM network while injecting trainable neural layers, significantly reducing the number of trainable parameters while maintaining performance on par with full fine-tuning. While PLMs combined with an upper-layer detection model perform well for English news, their effectiveness is limited in low-resource languages due to the lack of customized PLMs, as most PLMs are primarily developed for English. This raises the question of *how to develop effective misinformation detection in low-resource languages*. In low-resource languages, we lack well-trained LLMs; thus, approaches such as translation, etc. [6], have been employed. Our recent survey (status still under review), provides a detailed review of related work in low-resource languages and the gaps in the current literature. We explored a summarization-based approach [6], in which news content was first summarized and then analyzed for detection. This approach demonstrated strong performance despite concerns about potential information loss, highlighting the need for further investigation.

Fake news lacks a precise, universally accepted definition, often overlapping with terms like misinformation and rumor. While fact-checking methods exist, they fail to distinguish between inaccurate statements and intentional deception. As highlighted in our recent survey [8], we argue that fake news should be defined by its intent to mislead rather than factual inaccuracy alone. However, classifying intent remains challenging due to the absence of labeled datasets and clear evaluation metrics. We propose that user interactions—such as reposts, comments, and discussions—can serve as indicators of misleading intent in fake news. In our previous experiments [3], [5] and [7], we have integrated user comments and behavior patterns with news content for detection. Indeed, the results indicated that these additional information enhances detection performance, providing empirical support for our hypothesis. Building on this, we plan to develop a multi-module cross-attention framework that leverages Graph Neural Networks (GNNs) to model user interactions.

Another key question that arises, based on a thorough review of the related work, is *how to address the problem of domain shifting in FND*. Domain shifting occurs when a model trained on one domain (e.g., political news) is applied to another (e.g., health information). The feature distributions (e.g., language patterns, writing style) differ between training and target data, reducing model generalization across topics, platforms, or languages. To initially tackle this, we introduced a cross-domain few-shot learning approach [4] that excels in classification and transfer learning for FND. As a future direction, we are working on developing an effective DL model that leverages a Mixture of Experts approach to address this issue. In my final year, I will finish the Mixture of Experts approach and experiments by April 2025, focus on experimenting with LLM models from May to June, and implement the multi-module cross-attention approach by October. I will reserve three months for thesis writing, completing it by November 2025.

References

1. Towards covid-19 fake news detection using transformer-based models. Knowledge-Based Systems **274**, 110642 (2023)
2. Alghamdi, J., Lin, Y., Luo, S.: A comparative study of machine learning and deep learning techniques for fake news detection. Information **13**(12), 576 (2022)
3. Alghamdi, J., Lin, Y., Luo, S.: Does context matter? effective deep learning approaches to curb fake news dissemination on social media. Applied Sciences **13**(5), 3345 (2023)
4. Alghamdi, J., Lin, Y., Luo, S.: Cross-domain fake news detection using a prompt-based approach. Future Internet **16**(8), 286 (2024)
5. Alghamdi, J., Lin, Y., Luo, S.: Enhancing hierarchical attention networks with cnn and stylistic features for fake news detection. Expert Systems with Applications **257**, 125024 (2024)
6. Alghamdi, J., Lin, Y., Luo, S.: Fake news detection in low-resource languages: A novel hybrid summarization approach. Knowledge-Based Systems **296**, 111884 (2024)
7. Alghamdi, J., Lin, Y., Luo, S.: Unveiling the hidden patterns: A novel semantic deep learning approach to fake news detection on social media. Engineering Applications of Artificial Intelligence **137**, 109240 (2024)
8. Alghamdi, J., Luo, S., Lin, Y.: A comprehensive survey on machine learning approaches for fake news detection. Multimedia Tools and Applications **83**(17), 51009–51067 (2024)