Proceedings of 3rd International Conference on Interdisciplinary Humanities and Communication Studies
DOI: 10.54254/2753-7064/42/20242585

Exploring the Impact of Different Textual Language Features on Large Language Models' Detection of Fake News

Huilin Ouyang[1,a,*], Hangqi Yan[1]

[1]*School of Education, Soochow University, Suzhou, Jiangsu, 215000, China*
a. k24006011@kcl.ac.uk
**corresponding author*

Abstract: With the proliferation of social media and online platforms, fake news has become increasingly rampant. This study explores the impact of different textual language features on large language models (such as ChatGPT) in detecting fake news. By extracting extreme emotional vocabulary and exaggerated syntactic words commonly found in fake news and calculating their TF-IDF values, this study analyzes their influence on large language models' ability to assess the veracity of news. The study found that the frequency of extreme emotional words is higher than that of exaggerated syntactic words and has a more significant impact on fake news detection by large language models. Furthermore, this study suggests that by carefully selecting and adjusting language features, the accuracy and stability of fake news detection can be improved, providing new insights for optimizing automated detection systems. These findings provide important references for improving the technology of automatic fake news detection, contributing to the construction of a safer and more reliable online environment.

Keywords: Fake News Detection, Large Language Models, Language Features.

1. Introduction

With the rapid development of information technology, the Internet has become a primary channel for information dissemination. However, in the digital age, the widespread use of social media and online platforms has exacerbated the proliferation of fake news. These fake news items, often in the form of false or misleading information, are disguised as credible news reports. On a personal level, for example, a fake news story about Obama being injured in an explosion once led to a stock market crash, causing significant financial losses for many individuals. At the societal level, during natural disasters, fake news often emerges, causing public panic, such as during the 2011 Japan earthquake [1] and Hurricane Sandy in 2012 [2]. At the national level, the 2016 U.S. presidential election saw the spread of a large amount of fake news, seriously affecting the election results. These phenomena pose serious threats to public welfare, social trust, and democratic processes, potentially leading to panic, manipulation of public opinion, and influencing critical decisions. Therefore, timely detection of fake news, especially before it spreads widely on social networks, is crucial.

However, traditional manual review methods are not only inefficient but also struggle to cope with the challenge of vast amounts of information. Exploring accurate and efficient techniques for

DOI: 10.54254/2753-7064/42/20242585

detecting fake news holds great theoretical value and practical significance. Large language models (LLMs), such as the GPT series and BERT, have demonstrated strong text comprehension and generation capabilities in the field of natural language processing, offering new possibilities for the automatic detection of fake news. Additionally, different language features in the text significantly affect the accuracy and efficiency of LLMs in detecting fake news.

A large language model (LLM) is a critical technology in the field of natural language processing (NLP). It uses deep learning techniques to learn and understand the rules and patterns of natural language by analyzing massive amounts of text data. The operation of large language models is mainly divided into two stages: learning and prediction. The learning stage involves collecting large amounts of text data from various channels, including books, news, websites, and social media content, followed by preprocessing the data to filter out irrelevant information and retain valuable text data. Neural network algorithms are then used to train the processed text data, enabling the model to learn language rules and features such as vocabulary usage, sentence structure, and contextual meanings. The prediction stage occurs when the model receives a word or sentence and attempts to predict the next word, a process based on the language patterns and probability calculations learned by the model.

Currently, various methods have been applied to fake news detection. At its core, fake news detection is a classification problem, formally defined as follows: Given an event X and a series of news items related to X, denoted as M = {m1, m2, ..., mn}, these news items are published by a group of users U = {u1, u2, ..., un}. The news can take the form of social media content, such as posts on platforms like Weibo or Twitter, along with subsequent reposts, comments, or published articles and videos by the users. The goal of fake news detection is to learn a function f(x) that determines whether the event or a specific news item is false [3]. The basic paradigm of fake news detection can be divided into two steps: extracting relevant features and encoding these features for classification. In the first step, in addition to extracting common features such as semantic features, thematic features, sentiment features, visual features, temporal features, and user features, specific features can also be extracted based on different dissemination intents. In the second step, the extracted features are further encoded and used by classifiers to determine the authenticity of the news. Current classification methods can be categorized into traditional machine-learning methods and neural network-based methods. Traditional machine learning methods first select appropriate features and then use supervised or unsupervised methods for classification. Neural network-based methods can automatically select and integrate news features before classification and use convolutional neural networks and recurrent neural networks to process image features, text features, and temporal features in news content, allowing for multi-dimensional and multi-modal feature processing.

However, large language models (LLMs) still have certain limitations in processing complex texts. Specifically, with regard to ChatGPT, it may still struggle to correctly interpret ambiguous issues due to the lack of word segmentation, part-of-speech tagging, or syntactic analysis in some cases. For example, when asked to translate "无线电法国别研究" into English, ChatGPT misunderstood the meaning of the sentence due to incorrect word segmentation, leading to the erroneous translation as "Wireless French Research." Therefore, when faced with complex issues, ChatGPT could benefit from leveraging specialized models for structured prediction tasks, incorporating word segmentation, part-of-speech tagging, and syntactic analysis results as additional inputs, potentially improving its performance and stability.

This study, through an in-depth analysis of language features affecting large language models such as Wenxin Yiyan (Baidu's ERNIE) and ChatGPT in detecting fake news, aims to optimize model algorithms, improve detection accuracy and speed, and provide technical support for building a safer and healthier online environment. Additionally, it helps deepen our understanding of the relationship between language features and the laws of information dissemination.

DOI: 10.54254/2753-7064/42/20242585

2. Literature Review

As fake news continues to proliferate on the internet, many scholars have studied the challenges of detecting fake news and using large language models to identify it. According to research by Kai Shu and others, detecting fake news on social media is challenging due to the large, incomplete, unstructured, and noisy data generated by user interactions. They analyzed the characteristics of fake news from psychological and sociological perspectives, summarized existing fake news detection methods, including content-based and social context-based approaches, and discussed evaluation metrics and representative datasets. Additionally, they pointed out the open issues in the field of fake news detection and proposed future research directions, such as cross-media fake news detection, intention detection, model integration, and noise data handling [4].

Kai Shu and his colleagues also noted that fake news is deliberately created with subjective and provocative language for economic and political purposes, often including clickbait-style headlines. Linguistic features can capture different writing styles and sensational headlines to detect fake news. These features can be extracted at different levels of news articles, such as the character, word, sentence, and document levels. To capture the differences between fake and real news, existing studies use both general linguistic features and domain-specific linguistic features. General linguistic features commonly represent documents for various natural language processing tasks, such as 1) Lexical features, including character-level and word-level features, such as total word count, the number of characters per word, word frequency, and unique words; 2) Syntactic features, such as sentence-level features like the frequency of function words or phrases, or the frequency of punctuation and part-of-speech (POS) tags. Domain-specific linguistic features, often aligned with the news domain, may include cited words, external links, the number of images, and the average length of images. Additionally, deception detection features can also be extracted from writing styles to identify deceptive information for fake news detection [4].

Che Wanxiang and other scholars explored the challenges, opportunities, and development directions of natural language processing (NLP) in the era of large models. They analyzed the broad applications of large language models (LLMs) in NLP and their performance improvements, such as enhanced fluency and coherence in language generation and understanding. They emphasized that the core tasks in NLP in the era of large models are text classification, structured prediction, semantic analysis, and sentiment computation. Che pointed out that in statistical learning and early deep learning methods, traditionally structured prediction tasks like tokenization, part-of-speech tagging, and syntactic parsing serve to preprocess text, removing ambiguity and providing richer information for downstream tasks. However, these tasks also introduce noise due to cascading effects, affecting downstream applications [5].

Beizhe Hu and colleagues highlighted that fake news detection is a highly complex task because fake news often hides within complex clues and real contexts. Small language models (SLMs) perform poorly in this regard due to their limited knowledge and capabilities. In contrast, large language models (LLMs), such as GPT-3.5, have recently excelled in various tasks. Hu explored the potential of LLMs in fake news detection and proposed a new method that enhances SLMs' performance using LLMs, showing the potential of combining the two. They also suggested solutions tailored to different application scenarios [6].

Liu Hualing and colleagues provided a comprehensive review of fake news detection technologies. They summarized and classified the concepts related to multimodal fake news and analyzed the trends in single-modality and multimodality news datasets. They introduced single-modality fake news detection technologies based on machine learning and deep learning, which have been widely used in the field of fake news detection. However, due to the multimodal nature of fake news (e.g., text, images, videos), traditional single-modality methods cannot fully capture the deep logic of fake news,

DOI: 10.54254/2753-7064/42/20242585

making them ineffective in dealing with the challenges posed by multimodal fake news data. To address this, recent advances in multimodal fake news detection technologies were reviewed, with a focus on multi-stream architectures and graph architectures, exploring the underlying concepts and potential limitations of these approaches. Finally, they analyzed the current difficulties and bottlenecks in the field of fake news detection research and provided future research directions [7].

Despite the progress made in using large language models to detect fake news, in-depth research on how different textual language features (such as sentiment tendencies, word choice, and syntactic structures) specifically affect model performance remains insufficient [8]. These features may vary across different cultures and linguistic environments, and current research often overlooks these subtle differences.

3. Methodology

As research on fake news detection increases, researchers have developed numerous benchmark datasets, most of which are collected from real-life social media platforms like Twitter and Sina Weibo. This study uses existing fake news detection datasets, as shown in Table 1 below:

Table 1: Classic Data Sets in the Field of Fake News Detection

Name of the dataset	Total number of news articles	Total number of events	Number of fake news articles	type
BuzzFeedNews	1627	1627	901	text

The BuzzFeedNews dataset includes headlines and text of news stories from Facebook related to the 2016 U.S. election. It contains 1627 news items, of which 901 are fake news. This study used ChatGPT to identify fake news from this dataset and found that ChatGPT identified 1128 fake news items out of 1627, exceeding the original 901 fake news items in the dataset. This indicates that ChatGPT has a tendency to over-detect, mistakenly identifying 227 real news items as fake news.

The study then analyzed the 1128 news items identified (including the 227 incorrectly identified real news). According to Hassan et al., the TF-IDF (term frequency-inverse document frequency) value helps focus on words frequently used in fake news but rarely found in real news, such as "shocking" and "unimaginable" [9]. Building on the research by Chen et al. on extracting TF-IDF features [10], this study created a dictionary of keywords for all news items and classified these words based on linguistic features into "exaggerated syntactic expression words" and "extreme emotional expression words," calculating their TF-IDF values. "Exaggerated syntactic expression words" refer to sentences where subjects, predicates, and objects are exaggerated or attention-grabbing. For example, in "The star mysteriously disappeared," the exaggerated pairing of the subject "star" with the predicate "disappeared" would be recognized. "Extreme emotional expression words" refer to specific words and phrases that express extreme emotions, such as "terrifying," "stunning," and "deadly," or phrases like "catastrophic" and "devastating," which express strong negative emotions.

4. Results

Based on a detailed analysis of the TF-IDF values of each keyword in the dictionary, this study reveals a noteworthy phenomenon: in the group of words with a TF-IDF value exceeding 0.5, extreme emotional words account for a significant portion, reaching 57%. Although exaggerated syntactic expression words are also notable, they only account for 43%. This data distribution clearly shows that, in news texts, when the importance of words (measured by TF-IDF values) reaches a certain threshold, the occurrence frequency of extreme emotional words surpasses that of exaggerated syntactic expressions.

DOI: 10.54254/2753-7064/42/20242585

Upon further consideration of these findings, this study reasonably infers that when evaluating the ability of large language models to detect fake news, extreme emotional words appear to have a more prominent influence. These words are capable of strongly triggering emotional responses in readers, whether positive or negative, and can serve as a powerful tool for fake news creators to manipulate public opinion and mislead readers. In contrast, while exaggerated syntactic expressions can also attract attention, their direct impact on emotions may be slightly weaker. Therefore, in fake news detection mechanisms, extreme emotional words play a more critical role. In summary, extreme emotional words have a more significant influence on large language models in detecting fake news, emphasizing the indispensable role of sentiment analysis in enhancing the technology for verifying news authenticity.

5. Discussion

This study finds that in fake news, extreme emotional words occur more frequently than exaggerated syntactic words. Firstly, the higher frequency of extreme emotional words is partly due to their emotional appeal. Fake news or clickbait often uses extreme emotional words to grab readers' attention, such as "shocking," "astonishing," and "unbelievable." This approach can evoke emotional responses in readers, increasing click rates and the spread of the news. Therefore, extreme emotional words are commonly found in such content. Secondly, market-driven factors also contribute to the frequent use of emotional words. Many commercial or entertainment-oriented fake news stories are designed to elicit emotional resonance or provoke strong reactions, leading to the heavy use of emotional and exaggerated adjectives and adverbs. This makes emotional words much more prevalent in these types of articles.

On the other hand, the frequency of exaggerated syntactic words is relatively lower. One reason for this is that exaggerated syntax (such as "aliens disappeared") typically involves complex semantics or fictional scenarios, requiring careful design by the writer. The structure of such sentences is relatively unique and less common than the simple use of extreme emotional words. The exaggerated syntax is more commonly seen in specific types of articles (such as pseudoscience, conspiracy theories, or science fiction-themed fake news) and is not employed in all fake news. Additionally, exaggerated syntax is often used to describe fictional events or absurd scenarios, making it less universally present in all types of misinformation. Therefore, its usage frequency is lower, usually appearing only in certain themes (such as extraterrestrials, supernatural events, or unverified scientific discoveries).

Overall, extreme emotional words occur more frequently in fake news than exaggerated syntactic words because they can directly and quickly capture readers' emotional responses, while exaggerated syntax is more dependent on specific semantic contexts and fictional content.

The frequent occurrence of extreme emotional words has several effects on fake news detection. Firstly, these words possess distinctive features that can be quickly identified using existing natural language processing techniques (such as sentiment analysis tools or TF-IDF methods). When fake news frequently uses extreme emotional words, these words can serve as markers for misinformation, helping improve detection accuracy. Furthermore, the high frequency of these words provides a foundation for building automated detection systems. For example, fake news detection models can be programmed to flag texts containing extreme emotional words as high-risk, allowing for the rapid filtering of potential misinformation.

Secondly, because extreme emotional words often appear in fake news, the combination of these words with sentiment analysis allows models to detect such language more accurately, reducing the risk of misclassifying fake news as real information (i.e., reducing false negatives). Although extreme emotional words are an important feature, they work best when combined with other features, such as exaggerated syntax or dissemination patterns, to further improve the reliability of fake news

DOI: 10.54254/2753-7064/42/20242585

detection. A comprehensive analysis of multiple features can enhance the model's decision-making capability.

Thirdly, to avoid over-reliance on emotional words and reduce false positives, it is essential to recognize that real news may sometimes use extreme emotional words, particularly in reports of disastrous events, urgent warnings, or major breaking news. If the model depends too much on emotional words, it could mistakenly classify some real news as fake. Therefore, while emotional words are strong indicators of fake news, the model must also consider the context and other factors, such as news sources, timing, and user comments, to make balanced judgments. Additionally, the emotional words in fake news can vary in complexity, so the model needs to have a broad vocabulary to identify and interpret the semantic logic behind these words.

Finally, extreme emotional words are widely used in various types of fake news, such as commercial advertisements, political propaganda, and entertainment gossip. As a key feature, these words enable fake news detection models to generalize well across different fields of misinformation detection. Moreover, the model's sensitivity to emotional extremes, particularly in the news domain, allows it to handle different types of fake news more efficiently. This sensitivity to emotional content helps the model address new and emerging patterns of misinformation effectively.

6. Conclusion

This study, through an in-depth analysis of language features, reveals that the high frequency of extreme emotional words in fake news significantly impacts a large language model's ability to detect fake information. In comparison, exaggerated syntactic words are less frequent but still play an important role in specific contexts. The influence of extreme emotional words is more prominent, providing more reliable features for large language models to better detect fake news. However, relying too heavily on emotional words can also lead to the misclassification of some real news. Therefore, it is insufficient to rely solely on emotional features; other features, such as syntactic structure and contextual semantics, must also be considered to avoid both false positives and false negatives.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] *Takayasu, M., Sato, K., Sano, Y. (2015) Rumor diffusion and convergence during the 3.11 earthquake: A Twitter case study. PLoS One, 10(4): e0121443.*

[2] *Gupta, A., Lamba, H., Kumaraguru, P. (2013) Faking Sandy: Characterizing and identifying fake images on Twitter during Hurricane Sandy. Proceedings of the 22nd international conference on World Wide Web (pp. 729-736).*

[3] *Mao, Z., Zhao, B., Bai, J. (2022) A Review of Fake News Detection Methods Based on Propagation Intent Features. Signal Processing, 38(06): 1155-1169.*

[4] *Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H., (2021) Fake News Detection on Social Media: A Data Mining Perspective. Social and Information Networks.*

[5] *Che, W., Dou, Z., Feng, Y. (2023) Natural Language Processing in the Era of Large Models: Challenges, Opportunities, and Development. Science China: Information Sciences, 53(09): 1645-1687.*

[6] *Hu, B., Sheng, Q., Cao, J., Shi, Y., Li, Y., Wang, D., Qi, P. (2024) Bad Actor, Good Advisor: Exploring the Role of Large Language Models in Fake News Detection. Proceedings of the AAAI Conference on Artificial Intelligence, 38(20): 22105-22113.*

[7] *Liu, H., Chen, S., Cao, S. (2023) A Study on Fake News Detection Based on Multimodal Learning [J]. Computer Science and Exploration, 17(09): 2015-2029.*

[8] *Zhang, Z., Jing, J., Li, F. (2021), A Review of Fake Information Detection, Spread, and Control in Online Social Networks from the Perspective of Artificial Intelligence. Journal of Computer Science, 44(11): 2261-2282.*

DOI: 10.54254/2753-7064/42/20242585

[9] *Hassan, N., Li, C., Tremayne, M. (2015) Detecting Check-Worthy Factual Claims in Presidential Debates. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management.*
[10] *Chen, T., Li, X., Yin, H. (2018) Call Attention to Rumors: Deep Attention-Based Recurrent Neural Networks for Early Rumor Detection. Trends and Applications in Knowledge Discovery and Data Mining.*