JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

WELFake: Word Embedding Over Linguistic Features for Fake News Detection

Prof. P. A. Kharat 1) Rushikesh Madhukar Varade 2) Aniruddha Shrikrushna More 3) Gayatri Bhagwan Gond 4) Punam Ramesh Bhute.
Computer Science Engineering Department, Padm. Dr. VBKCOE, Malkapur, Maharashtra, India.

ABSTRACT *- The paper introduces WELFake, a novel approach for fake news detection that leverages word embeddings over linguistic features. WELFake combines the power of word embeddings with linguistic features to enhance the accuracy of fake news identification.*
The proposed methodology demonstrates the importance of linguistic nuances and context in distinguishing between genuine and deceptive information. By integrating word embeddings with linguistic features, WELFake aims to improve the robustness of fake news detection models. The experimental results showcase the effectiveness of WELFake in achieving higher accuracy and reliability compared to traditional methods.

Keywords *- Word Embedding, Linguistic Features, Fake News Detection, Natural Language Processing, Machine Learning.*

INTRODUCTION *- In today's digital age, the rapid dissemination of information through online platforms has given rise to the urgent need for reliable mechanisms to discern the authenticity of news articles. The introduction of fake news poses a significant threat to societal well-being and trust in media.[1] This paper addresses this critical issue by proposing WELFake, a sophisticated model that combines word embeddings and linguistic features for enhanced fake news detection. The integration of linguistic features allows WELFake to consider syntactic and semantic aspects, providing a more nuanced understanding of language, which is crucial for accurate classification.[2]*

The escalating challenge of fake news demands advanced solutions that can adapt to the evolving strategies employed by purveyors of deceptive information. WELFake not only contributes to the academic discourse on fake news detection but also provides a practical and effective tool to mitigate the societal impact of misinformation.[3]
The following sections delve into the specific problems addressed by WELFake, the proposed system methodology, and the underlying languages and technologies used in its development.[4]

PROBLEM FORMULATION *-*

The rampant spread of fake news on social media platforms and online news websites has raised concerns about the authenticity and reliability of information available to the public. Traditional methods of detecting fake news often rely on simplistic features such as headline characteristics or source credibility, which may not capture the nuanced linguistic patterns indicative of misinformation. In this paper, we address the challenge of fake news detection by formulating a novel approach that integrates word embedding over linguistic features.[5]

The primary objective of our study is to develop an effective and robust framework for identifying fake news articles based on their linguistic content. To achieve this, we aim to leverage the semantic and syntactic information embedded in text data using advanced natural language processing techniques. By combining word embedding models with linguistic feature extraction methods, we seek to enhance the accuracy and reliability of fake news detection across diverse contexts and domains.[6]

Our proposed methodology aims to overcome the limitations of existing fake news detection methods by providing a more comprehensive analysis of textual content. By capturing both semantic meanings and syntactic structures of news articles, we aim to uncover subtle linguistic cues indicative of fake news. Through rigorous experimentation and evaluation, we intend to demonstrate the effectiveness of our approach in accurately identifying fake news articles and mitigating the spread of misinformation in online platforms.[7]

PROPOSE SYSTEM METHODOLOGY -

WELFake introduces a comprehensive methodology for fake news detection, integrating word embeddings with linguistic features. The proposed system methodology is detailed in five key steps. Firstly, linguistic features are extracted from the text, capturing syntactic and semantic aspects. Then, word embeddings are generated
to represent the underlying context. [8] The fusion of linguistic features and word embeddings creates a robust representation of the text, allowing for a more nuanced analysis. The paper elaborates on the training and testing phases, highlighting the significance of leveraging both linguistic and contextual information for improved fake news detection. The proposed methodology is supported by experimental results that demonstrate its superiority over traditional approaches.[9]

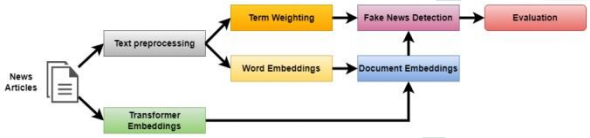

Fig. 1.1 Embedding And Working Methodology Of Fake News Detection

WORKING ON LANGUAGES -

Hardware Technology:

In recent years, hardware technology has witnessed significant advancements, driving innovation across various sectors and industries. The evolution of processors, such as the Pentium i3, has enabled faster computation and improved multitasking capabilities. Solid-state drives (SSDs) and high-capacity hard disks, like the 500 GB mentioned, have revolutionized storage solutions, offering faster data access and greater reliability. Moreover, advancements in display technology have led to the development of LED monitors, which provide clearer visuals and consume less power compared to traditional displays. Input devices like keyboards and mice have also evolved with ergonomic designs and enhanced functionalities, catering to user comfort and productivity. Additionally, the integration of higher RAM capacities, like 4 GB, ensures smoother multitasking and improved system performance, supporting resource-intensive applications and tasks.

Software Technology:

Software technology has undergone rapid transformation, with operating systems like Windows 10 offering enhanced security, improved user interfaces, and advanced features. Programming languages like Python have gained immense popularity due to their simplicity, versatility, and extensive libraries, facilitating the development of complex applications and solutions. Web frameworks such as Flask provide a streamlined approach to web development, enabling developers to build scalable and efficient web applications with ease. Moreover, advancements in artificial intelligence (AI) and machine learning algorithms have paved the way for innovative solutions in various domains, from healthcare to finance. Cloud computing platforms and services have also gained traction, offering scalable and cost-effective solutions for storage, computation, and data analytics, empowering organizations to harness the power of big data and AI-driven insights..

RELATED WORKING -

Fake news detection has garnered significant attention due to its detrimental effects on society, including misinformation dissemination and manipulation of public opinion. Various approaches have been proposed to tackle this issue, leveraging techniques from natural language processing (NLP), machine learning (ML), and deep learning (DL).[10] One notable approach involves the utilization of linguistic features combined with word embeddings for more effective fake news detection. Linguistic features encompass syntactic and semantic characteristics of text, such as grammatical structures, lexical choices, and sentiment. Integrating these features with word embeddings, which capture semantic similarities between words in high-dimensional vector spaces, can enhance the understanding of textual content and improve the accuracy of fake news detection models.[11]
Another line of research in fake news detection explores the use of ensemble methods, which combine multiple models to achieve better performance than individual classifiers. Ensemble methods often involve the aggregation of predictions from diverse models, such as decision trees, support vector machines, and neural networks, to mitigate biases and increase robustness. Additionally, deep learning architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promise in capturing intricate patterns in textual data for fake news detection. These models leverage hierarchical feature representations and sequential dependencies to extract nuanced information from news articles, tweets, and other textual sources.[12]

Furthermore, domain-specific knowledge and expertise play a crucial role in fake news detection. Researchers have developed specialized datasets and ontologies tailored to the characteristics of fake news, enabling more targeted analysis and model development. Additionally, techniques such as transfer learning, which leverages knowledge gained from pre-trained models on large-scale corpora, have been adapted to the fake news detection domain. By fine-tuning pre-trained models on task-specific datasets, transfer learning facilitates the transfer of knowledge from related tasks and domains, thereby enhancing the generalization and adaptability of fake news

detection systems across different contexts and languages.[13]

HARDWARE TECHNOLOGY -

System: Pentium i3 Processor.
Hard Disk: 500 GB.
Monitor: 15'' LED.
Input Devices: Keyboard, Mouse.
RAM: 4 GB.

SOFTWARE TECHNOLOGY -

Operating System: Windows 10.
Coding Language: Python.
Web Framework: Flask.

REFERENCES -

[1] W. Jiang, J. Wu, F. Li, G. Wang, and H. Zheng, "Trust evaluation in online social networks using generalized network flow," IEEE Trans. Comput., vol. 65, no. 3, pp. 952–963, Mar. 2016.

[2] M. Alrubaian, M. Al-Qurishi, A. Alamri, M. Al-Rakhami, M. M. Hassan, and G. Fortino, "Credibility in online social networks: A survey," IEEE Access, vol. 7, pp. 2828–2855, 2019.

[3] S. Ranganath, S. Wang, X. Hu, J. Tang, and H. Liu, "Facilitating time critical information seeking in social media," IEEE Trans. Knowl. Data Eng., vol. 29, no. 10, pp. 2197–2209, Oct. 2017.

[4] Z. Zhang, R. Sun, X. Wang, and C. Zhao, "A situational analytic method for user behavior pattern in multimedia social networks," IEEE Trans. Big Data, vol. 5, no. 4, pp. 520–528, Dec. 2019.

[5] M. Schudson and B. Zelizer, "Fake news in context," in Understanding and Addressing the Disinformation Ecosystem. Philadelphia, PA, USA: Annenberg School for Communication, Apr. 2017, pp. 1–4.

[6] S. Zaryan, "Truth and trust: How audiences are making sense of fake news," M.S. thesis, Media Commun. Studies, Lund Univ. Publications Student Papers, Stockholm, Sweden, Jun. 2017. [Online]. Available: https://lup.lub.lu.se/student-papers/search/publication/8906886

[7] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018.

[8] R. Sequeira, A. Gayen, N. Ganguly, S. K. Dandapat, and J. Chandra, "A large-scale study of the Twitter follower network to characterize the spread of prescription drug abuse tweets," IEEE Trans. Comput. Social Syst., vol. 6, no. 6, pp. 1232–1244, Dec. 2019.

[9] Politifact News Dataset. Accessed: Mar. 31, 2020. [Online]. Available: http://www.politifact.com/

[10] The Washingtonpost Fact Checker. Accessed: Mar. 31, 2020. [Online]. Available: https://www.washingtonpost.com/news/fact-checker.

[11] Fact Check. Accessed: Mar. 31, 2020. [Online]. Available: https://www. factcheck.org/

[12] Snopes. Accessed: Mar. 31, 2020. [Online]. Available: https://www. snopes.com/

[13] Truthorfiction. Accessed: Mar. 31, 2020. [Online]. Available: https:// www.truthorfiction.com/
[14] Fullfact. Accessed: Mar. 31, 2020. [Online]. Available: https://fullfact. org/

[15] Hoax Slayer. Accessed: Mar. 31, 2020. [Online]. Available: http://hoaxslayer.com/

[16] Viswas News. Accessed: Mar. 31, 2020. [Online]. Available: http://www. vishvasnews.com/

[17] Factly. Accessed: Mar. 31, 2020. [Online]. Available: https://factly.in/

[18] L.-L. Shi et al., "Human-centric cyber social computing model for hotevent detection and propagation," IEEE Trans. Comput. Social Syst., vol. 6, no. 5, pp. 1042–1050, Oct. 2019.

[19] M. Glenski, T. Weninger, and S. Volkova, "Propagation from deceptive news sources who shares, how much, how evenly, and how quickly?" IEEE Trans. Comput. Social Syst., vol. 5, no. 4, pp. 1071–1082, Dec. 2018.

[20] E. Lancaster, T. Chakraborty, and V. S. Subrahmanian, "MALT P : Parallel prediction of malicious tweets," IEEE Trans. Comput. Social Syst., vol. 5, no. 4, pp. 1096–1108, Dec. 2018.

[21] P. K. Verma and P. Agrawal, "Study and detection of fake news: P2C2-based machine learning approach," in Proc. Int. Conf. Data Manage., Anal. Innov., vol. 1175. Singapore: Springer, Sep. 2020, pp. 261–278.

[22] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," IEEE Trans. Multimedia, vol. 19, no. 3, pp. 598–608, Mar. 2017.

[23] B. Ratner, "The correlation coefficient: Its values range between +1/ − 1, or do they?" J. Targeting, Meas. Anal. Marketing, vol. 17, no. 2, pp. 139–142, Jun. 2009.

[24] A. De Salve, P. Mori, B. Guidi, and L. Ricci, "An analysis of the internal organization of Facebook groups," IEEE Trans. Comput. Social Syst., vol. 6, no. 6, pp. 1245–1256, Dec. 2019.

[25] P. K. Verma, P. Agrawal, and R. Prodan, WELFake Dataset for Fake News Detection in Text Data (Version: 0.1) [Data Set]. Genéve, Switzerland: Zenodo, 2021.