10.48047/jocaaa.2024.32.02.37

CLASSIFICATION OF FAKE NEWS UTILISING TENSOR DECOMPOSITION AND GRAPH CONVOLUTIONAL NETWORK

Mrs.K.RAJANI[1] , KONGARA SINDHU[2], JOGU SHRUTHI[3], MERUGU PRANAVI[4]

[1]Assistant Professor, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India

Email: swecrajani86@gmail.com

[2,3,4,]B.Tech Student, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India

Abstract: With the wide spread of fake news on social media, its impact has become a major concern of the public, so accurate detection methods are urgently needed. However, these methods rarely investigate the sentence interaction patterns of different news articles, and most of them do not consider fine-grained fake news classification. To overcome these issues, we propose a method that constructs a graph representation for news articles and employs a graph neural network (GNN) to classify fake news. The method uses the local word co-occurrence information of sentences to obtain the interaction relationship between sentences, which is abstracted by the weight matrix of the graph representation. A third-order co-occurrence tensor is built, and the weight matrix is calculated based on the canonical polyadic (CP) decomposition of this tensor. The computed representations can capture more accurate contextual information of news articles. The results on two real-world datasets demonstrate that the proposed method outperforms the competing methods in both binary and multiclass classification tasks. In particular, for multiclass classification on the selected dataset with 70% of the training set for training, the improvements got good accuracy and F1-score.

***Index Terms** - Fake news, Social media, Sentence interaction patterns, Graph representation, Graph neural network (GNN), Local word co-occurrence, Third-order co-occurrence tensor, Canonical polyadic (CP) decomposition, Contextual information, Multiclass classification.*

Keywords: Fake news, Social media, Sentence interaction patterns, Graph representation, Graph neural network (GNN), Local word co-occurrence, Third-order co-occurrence tensor, Canonical polyadic (CP) decomposition, Contextual information, Multiclass classification.

1. INTRODUCTION

In the digital age, the rampant spread of fake news through social media platforms has emerged as a pressing societal concern. Detecting and classifying fake news accurately is a vital endeavor, yet existing methods often fall short in analyzing sentence interaction patterns within news articles and fail to provide fine-grained classification. To address these challenges, our project introduces an innovative approach that constructs a graph-based representation for news articles and utilizes Graph Neural Networks

10.48047/jocaaa.2024.32.02.37

(GNNs) for classification. By leveraging local word co-occurrence data and a unique tensor-based approach, our method enhances contextual understanding and out performs existing methods, as evidenced by substantial improvements in accuracy and F1-score on real-world datasets. This introduction sets the stage for our project's contribution to combating the proliferation of fake news in the digital landscape.

The proliferation of fake news has become a critical challenge in the era of information overload, particularly evident during significant events such as the 2016 US presidential election. Scholars and researchers have engaged in a comprehensive exploration of this phenomenon, employing diverse methodologies to understand, detect, and mitigate the dissemination of misinformation. Notable studies, such as Grinberg et al. [1], Bondielli and Marcelloni [2], and Zhou and Zafarani [3], have delved into the landscape of fake news, offering insights into its characteristics and the techniques for detection.

Machine learning paradigms have played a pivotal role in tackling this issue, as evidenced by works like Shahid et al.'s exploration of challenges and future research opportunities in detecting and mitigating fake news [4]. Various approaches, from ensemble learning models [7] to deep contextualized text representation [6], reflect the dynamism of the field. Additionally, researchers have explored the integration of domain-specific features [16], early detection strategies [17], and the impact of social context [18] on the credibility assessment of news.

This introduction surveys seminal contributions in the domain, encompassing tensor factorization [13], graph neural networks [12], and hybrid deep models [20]. The multifaceted nature of the research underscores the urgency of addressing fake news through interdisciplinary and evolving methodologies. As the information landscape continues to evolve, this body of work provides a foundation for ongoing efforts to combat the challenges posed by fake news.

2. LITERATURE REVIEW

Fake news detection has become a crucial research area, particularly in the context of the information age and the digital era where misinformation can spread rapidly. Numerous studies have been conducted to understand, characterize, and develop techniques for identifying and mitigating the impact of fake news. This literature survey provides an overview of key research papers in this field.

Grinberg et al. [1] conducted a study on fake news during the 2016 US presidential election on Twitter, highlighting the prevalence and impact of misinformation in a real-world context. Bondielli and Marcelloni [2] presented a comprehensive survey on fake news and rumor detection techniques, discussing various approaches and challenges. Zhou and Zafarani [3] further extended this survey, providing insights into fundamental theories, detection methods, and research opportunities in the field.

Shahid et al. [4] discussed challenges and future research opportunities in detecting and mitigating the dissemination of fake news, emphasizing the need for robust methods. Zhang and Ghorbani [5] offered an overview of online fake news, covering its characterization, detection, and broader discussion. The survey by Samadi et al. [6] focused on deep contextualized text representation and learning for

10.48047/jocaaa.2024.32.02.37

fake news detection, exploring advanced techniques for improved accuracy.

Several studies have investigated machine learning paradigms for fake news detection. Huang and Chen [7] proposed an ensemble learning model based on self-adaptive harmony search algorithms. Katsaros et al. [8] addressed the question of which machine learning paradigm is most effective for fake news detection. Vaibhav et al. [9] explored the significance of sentence interactions in leveraging sentence-level representations for fake news classification.

Hosseinimotlagh and Papalexakis [10] presented an unsupervised content-based identification of fake news articles using tensor decomposition ensembles, introducing a novel approach. Wang [11] introduced a benchmark dataset named "Liar, liar pants on fire" for fake news detection, providing a standardized evaluation framework for researchers.

Graph-based methods have also been employed for fake news detection. Song, Shu, and Wu [12] proposed a temporally evolving graph neural network, demonstrating the effectiveness of graph-based approaches in capturing evolving patterns. Papanastasiou, Katsimpras, and Paliouras [13] utilized tensor factorization with label information for fake news detection, contributing to the diversity of detection techniques.

Network-based approaches have gained attention as well. Zhou and Zafarani [14] proposed a pattern-driven approach for network-based fake news detection, focusing on the structural aspects of information dissemination. Jin et al. [15] addressed news verification by exploiting conflicting social viewpoints in microblogs, highlighting the importance of social context.

Silva et al. [16] embraced domain differences in fake news detection, introducing a cross-domain approach using multimodal data. Yuan et al. [17] explored early detection of fake news by leveraging the credibility of news, publishers, and users based on weakly supervised learning.

Shu, Wang, and Liu [18] went beyond news contents, emphasizing the role of social context for fake news detection. Reis et al. [19] focused on supervised learning for fake news detection, providing insights into the effectiveness of machine learning models in this domain. Ruchansky, Seo, and Liu [20] introduced a hybrid deep model called CSI for fake news detection, demonstrating the potential of combining deep learning techniques.

In conclusion, the literature on fake news detection is diverse, encompassing various methodologies such as machine learning, graph-based methods, and network-based approaches. Researchers have explored different aspects, including text representation, social context, and domain adaptation, to enhance the accuracy and robustness of fake news detection systems. The challenges identified in these studies pave the way for future research directions in this critical area

3. METHODOLOGY

In previous research they focus on unsupervised fake news detection. They use tensor decompositions to capture the latent relations between articles and terms. Furthermore, an ensemble method is introduced to leverage multiple tensor

10.48047/jocaaa.2024.32.02.37

decompositions into a single, high-quality, and high-coherences to article clusters. In another research they focus on detecting fake news solely based on the structural information of social networks. They suggest that the underlying network connections of users that share fake news are discriminative enough to support the detection of fake news. They model each post as a network of friendship interactions and represent a collection of posts as a multidimensional tensor. Taking into account the available labeled data, they developed a tensor factorization method which associates the class labels of data samples with their latent representations. Specifically, they combine a classification error term with the standard factorization in a unified optimization process.

Drawbacks:

1. The existing work focuses on unsupervised fake news detection using tensor decompositions. This approach might require a large amount of unlabeled data, making it potentially less effective in cases where labeled data is limited.

2. The existing work relies on tensor decompositions to capture latent relations between articles and terms, which may have limitations in effectively representing complex relationships within the data.

3. Tensor decompositions used in the existing work might become computationally expensive as the dataset grows larger.

4. The ensemble method used in the existing work might provide good clustering performance, but it might be challenging to interpret and understand the decisions made by the model.

We propose a method that constructs a graph representation for news articles and employs a graph neural network (GNN) to classify fake news. The method uses the local word co-occurrence information of sentences to obtain the interaction relationship between sentences, which is abstracted by the weight matrix of the graph representation. Our framework consists of three main modules. The first one is a module for article graph representation, where an LSTM model is employed to obtain the feature vector of each vertex, and a CP decomposition-based method is used to obtain the weight matrix. The second module is designed to obtain the feature vector of the article by using the graph convolutional layer and the max pooling operation. The last one is the classification module, where a fully connected layer is adopted.

Benefits:

1. The present work, on the other hand, adopts a supervised approach using a GNN for classification, which can leverage labeled data and potentially achieve higher accuracy.

2. This graph-based representation can potentially capture more nuanced and meaningful features for fake news detection.

3. The GNN-based approach in the present work operates on the graph representation, which allows for better interpretability, as the relationships between sentences and their

10.48047/jocaaa.2024.32.02.37

weights in the graph can be analyzed to understand how fake news is classified.

4. By employing an LSTM model for obtaining the feature vector of each vertex (sentence)and using CP decomposition-based method to obtain the weight matrix, the model can extract high-level abstractions from the articles, enabling better discrimination between fake and legitimate news.

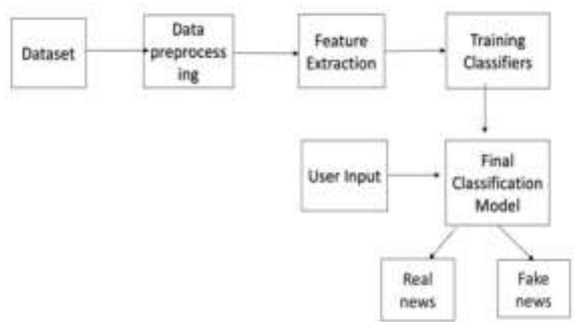

Fig 1 System Architecture

Modules:

The modules are:

- Data loading: using this module we are going to import the dataset.
- Data Preprocessing: using this module we will explore the data.
- Splitting data into train & test: using this module data will be divided into train & test
- Model generation: Model building - SVM - LSTM -CNN -GNN -BERT GCN - GCN with CP (Pytorch Graph CNN) -BERT GCN + LSTM -LSTM + GRU. Algorithms accuracy calculated
- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed

4. IMPLEMENTATION

SVM - SVM is a powerful supervised algorithm that works best on smaller datasets but on complex ones. Support Vector Machine, abbreviated as SVM can be used for both regression and classification tasks, but generally, they work best in classification problems.

LSTM – It is a special type of Recurrent Neural Network which is capable of handling the vanishing gradient problem faced by RNN. LSTM was designed by Hochreiter and Schmidhuber that resolve the problem caused by traditional RNNs and machine learning algorithms. LSTM can be implemented in Python using the Keras library.

CNN– A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data. There are other types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice.

GNN – Graph Neural Networks (GNNs) are a class of deep learning methods designed to perform inference on data described by graphs. GNNs are neural networks that can be directly applied to

10.48047/jocaaa.2024.32.02.37

graphs, and provide an easy way to do node-level, edge-level, and graph-level prediction tasks.

BERT GCN– BERT uses a multi-layer bidirectional transformer encoder to represent the input text in a high-dimensional space. That means it can take into account the entire context of each word in the sentence, which helps it to better understand the meaning of the text. The GCN is used to capture the spatial correlation between the nodes.

GCN with CP (Pytorch Graph CNN) – The graph convolutional network (GCN) was first introduced by Thomas Kipf and Max Welling in 2017. A GCN layer defines a first-order approximation of a localized spectralilter on graphs. GCNs can be understood as a generalization of convolutional neural networks to graph-structured data.

BERT GCN + LSTM – A Graph Convolutional Network, or GCN, is an approach for semi-supervised learning on graph-structured data. It is based on an efficient variant of convolutional neural networks which operate directly on graphs.

LSTM + GRU- Gated recurrent units (GRUs) are a gating mechanism in recurrent neural networks, introduced in 2014 by Kyunghyun Cho et al. The GRU is like a long short-term memory (LSTM) with a gating mechanism to input or forget certain features, but lacks a context vector or output gate, resulting in fewer parameters than LSTM.

5. EXPERIMENTAL RESULTS

Dataset Description:

ISOT Fake News Dataset

The dataset contains two types of articles fake and real News. This dataset was collected from realworld sources; the truthful articles were obtained by crawling articles from Reuters.com (News website). As for the fake news articles, they were collected from different sources. The fake news articles were collected from unreliable websites that were flagged by Politifact (a fact-checking organization in the USA) and Wikipedia. The dataset contains different types of articles on different topics, however, the majority of articles focus on political and World news topics.

The dataset consists of two CSV files. The first file named “True.csv” contains more than 12,600 articles from reuter.com. The second file named “Fake.csv” contains more than 12,600 articles from different fake news outlet resources. Each article contains the following information: article title, text, type and the date the article was published on. To match the fake news data collected for kaggle.com, we focused mostly on collecting articles from 2016 to 2017. The data collected were cleaned and processed; however, the punctuations and mistakes that existed in the fake news were kept in the text.

News	**Size (Number of articles)**	***Subjects***	
Real-News	21417	**Type**	**Articles size**
		World-News	*10145*
		Politics-News	*11272*
Fake-News	23481	**Type**	**Articles size**
		Government-News	*1570*
		Middle-east	*778*
		US News	*783*
		left-news	*4459*
		politics	*6841*
		News	*9050*

10.48047/jocaaa.2024.32.02.37

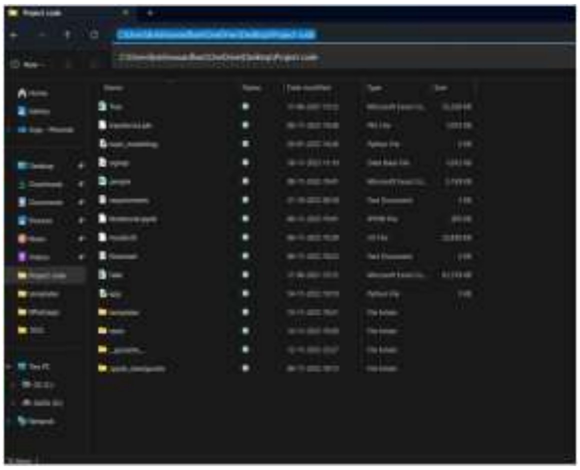

Fig 2 Project Code Folder

Fig 3 Anaconda prompt

Fig 4 URL

Fig 5 Link for website

Fig 6 User home page

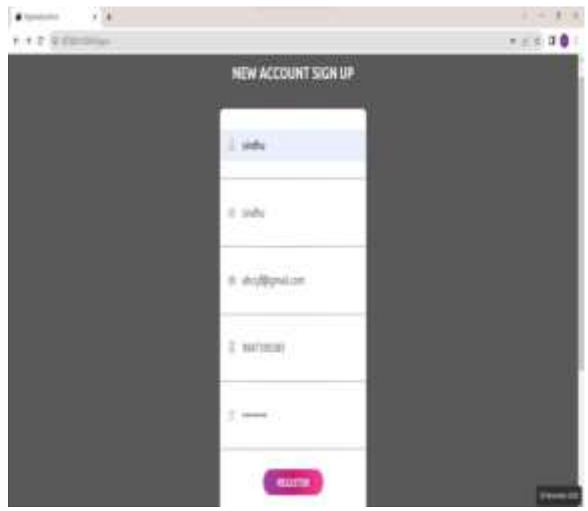

Fig 7 Signup Credentials

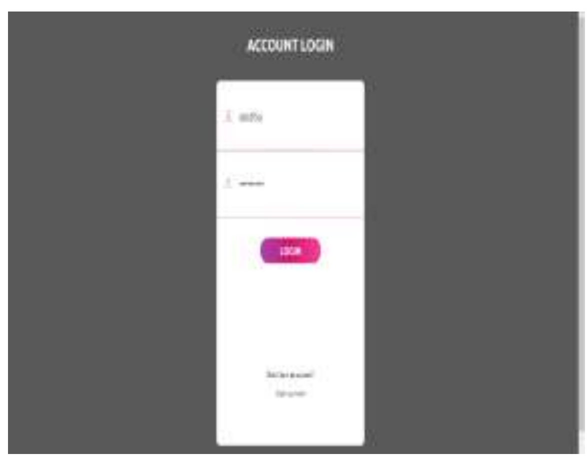

Fig 8 Login Credentials

10.48047/jocaaa.2024.32.02.37

Fig 9 Home page

Fig 10 InputForm

Fig 11 Outputscreen

6. CONCLUSION

This work focuses on fake news classification using only the body text of news. The key idea is to investigate the intra-article interaction of sentences for improvement. An effective approach using tensor decomposition and GCN is proposed, in which local lexical sentence-wise structures and sequential patterns are considered simultaneously. Specifically, we represent each article as a weighted graph whose vertices are the sentences of the article. To capture the semantic information in the graph representation, an LSTM is employed to compute the feature vectors of vertices. On the other hand, to capture local structural information between words for quantifying sentence interactions, CP decomposition based method is presented to compute the weight matrix. Furthermore, the graph is input into a GCN for classification, where the interaction between sentences can be captured. Since the local lexical structures and sequential patterns are jointly investigated, we can produce discriminative representations for news articles, thereby achieving better classification performance. Experimental results obtained on two real-world benchmark datasets for fake news classification show that the proposed method outperforms the competing methods and the generalizability of our method is illustrated.

FUTUREENHANCEMENTS

In future advancements for fake news classification using tensor decomposition and Convolutional Neural Networks (CNNs), our project could be to incorporate multi modal information. Instead of solely relying on textual data, consider integrating

10.48047/jocaaa.2024.32.02.37

other modalities like images, videos, or user engagement metrics (e.g., comments, shares).

REFERENCES

[1] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, "Fake news on Twitter during the 2016 US presidential election," Science, vol. 363, no. 6425, pp. 374–378, 2019.

[2] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," Inf. Sci., vol. 497, pp. 38–55, Sep. 2019.

[3] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, 2020.

[4] W. Shahid et al., "Detecting and mitigating the dissemination of fake news: Challenges and future research opportunities," IEEE Trans. Computat. Social Syst., early access, Jun. 6, 2022, doi: 10.1109/TCSS.2022.3177359.

[5] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," Inf. Process. Manag., vol. 57, no. 2, Mar. 2020, Art. no. 102025.

[6] M. Samadi, M. Mousavian, and S. Momtazi, "Deep contextualized text representation and learning for fake news detection," Inf. Process. Manag., vol. 58, no. 6, Nov. 2021, Art. no. 102723.

[7] Y.-F. Huang and P.-H. Chen, "Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms," Exp. Syst. Appl., vol. 159, Nov. 2020, Art. no. 113584.

[8] D. Katsaros, G. Stavropoulos, and D. Papakostas, "Which machine learning paradigm for fake news detection?" in Proc. IEEE/WIC/ACM Int. Conf. Web Intell., Oct. 2019, pp. 383–387.

[9] V. Vaibhav, R. Mandyam, and E. Hovy, "Do sentence interactions matter? Leveraging sentence level representations for fake news classification," in Proc. 13th Workshop Graph-Based Methods Natural Lang. Process., 2019, pp. 134–139.

[10] S. Hosseinimotlagh and E. E. Papalexakis, "Unsupervised contentbased identification of fake news articles with tensor decomposition ensembles," in Proc. Workshop Misinformation Misbehavior Mining Web, 2018, pp. 1–8.

[11] W. Y. Wang, "'Liar, liar pants on fire': A new benchmark dataset for fake news detection," in Proc. 55th Annu. Meeting Assoc. Comput. Linguistics, 2017, pp. 422–426.

[12] C. Song, K. Shu, and B. Wu, "Temporally evolving graph neural network for fake news detection," Inf. Process. Manag., vol. 58, no. 6, Nov. 2021, Art. no. 102712.

[13] F. Papanastasiou, G. Katsimpras, and G. Paliouras, "Tensor factorization with label information for fake news detection," in Proc. Int. Workshop Reducing Online Misinformation Exposure, Paris, France, Jul. 2019, pp. 1–7.

[14] X. Zhou and R. Zafarani, "Network-based fake news detection: A pattern-driven approach," ACM SIGKDD Explor. Newslett., vol. 21, no. 2, pp. 48–60, 2019.

10.48047/jocaaa.2024.32.02.37

[15] Z. Jin, J. Cao, Y. Zhang, and J. Luo, “News verification by exploiting conflicting social viewpoints in microblogs,” in Proc. AAAI Conf. Artif. Intell., 2016, pp. 2972–2978.

[16] A. Silva, L. Luo, S. Karunasekera, and C. Leckie, “Embracing domain differences in fake news: Cross-domain fake news detection using multimodal data,” in Proc. AAAI Conf. Artif. Intell., 2021, vol. 35, no. 1, pp. 557–565.

[17] C. Yuan, Q. Ma, W. Zhou, J. Han, and S. Hu, “Early detection of fake news by utilizing the credibility of news, publishers, and users based on weakly supervised learning,” in Proc. 28th Int. Conf. Comput. Linguistics, 2020, pp. 5444–5454.

[18] K. Shu, S. Wang, and H. Liu, “Beyond news contents: The role of social context for fake news detection,” in Proc. 12th ACM Int. Conf. Web Search Data Mining, Jan. 2019, pp. 312–320.

[19] J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, “Supervised learning for fake news detection,” IEEE Intell. Syst., vol. 34, no. 2, pp. 76–81, Mar./Apr. 2019.

[20] N. Ruchansky, S. Seo, and Y. Liu, “CSI: A hybrid deep model for fake news detection,” in Proc. ACM Conf. Inf. Knowl. Manag., Nov. 2017, pp. 797–806.