

Research Article

Multimodal Fake News Detection Incorporating External Knowledge and User Interaction Feature

Lifang Fu ¹ and Shuai Liu ²

¹College of Letters and Science, Northeast Agricultural University, Harbin 150000, China

²College of Engineering, Northeast Agricultural University, Harbin 150000, China

Correspondence should be addressed to Lifang Fu; lifangfu@neau.edu.cn

Received 8 December 2022; Revised 1 April 2023; Accepted 12 April 2023; Published 3 July 2023

Academic Editor: Zhongxu Hu

Copyright © 2023 Lifang Fu and Shuai Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of online social media, the number of various news has exploded. While social media provides an information platform for news release and dissemination, it also makes fake news proliferate, which may cause potential social risks. How to detect fake news quickly and accurately is a difficult task. The multimodal fusion fake news detection model is the current research focus and development trend. However, in terms of content, most existing methods lack the mining of background knowledge hidden in the news content and ignore the connection between background knowledge and existing knowledge system. In terms of the propagation chain, the research tends to emphasize only the single chain from the previous communication node, ignoring the intricate communication chain and the mutual influence relationship among users. To address these problems, this paper proposes a multimodal fake news detection model, A-KWGCN, based on knowledge graph and weighted graph convolutional network (GCN). The model fully extracted the features of the content and the interaction between users of the news dissemination. On the one hand, the model mines relevant knowledge concepts from the news content and links them with the knowledge entities in the wiki knowledge graph, and integrates knowledge entities and entity context as auxiliary information. On the other hand, inspired by the “similarity effect” in social psychology, this paper constructs a user interaction network and defines the weighted GCN by calculating the feature similarity among users to analyze the mutual influence of users. Two public datasets, Twitter15 and Twitter16, are selected to evaluate the model, and the accuracy reaches 0.905 and 0.930, respectively. In the comparison experiments, A-KWGCN model has more significant advantages than the other six comparison models in four evaluation indexes. Also, ablation experiments are conducted to verify that knowledge module and weighted GCN module play the significant role in the detection of fake news.

1. Introduction

News is information reported by mass media and based on objective facts. With the development of the Internet, online social media has gradually become the mainstream platform for publishing and disseminating news. The booming development of social media has led to an exponential growth of news. On the one hand, the rapid growth of news can bring convenience. On the other hand, most of the news is not vetted for authenticity before publication, and consequently, the problem of the proliferation of fake news arises.

Any inaccurate information that deviates from objective reality can be called fake news [1]. In the Internet era, fake

news spreads faster and has a worse impact. If fake news is not curbed, it will not only damage the credibility of the media but also bring serious harm to the society. Therefore, it becomes important to accurately and effectively identify the veracity of news and stop the spread of fake news.

However, fake news detection is a challenging task. The difficulties comprise four major components: (1) fake news is confusing and misleading in expression, and it is difficult to determine the veracity of news merely based on the analysis of news content. (2) News usually uses short texts to describe events, which makes it necessary to analyze the news content from various perspectives to discover information and features. (3) The dissemination process of fake news is

complex, and there is no rule or formula to describe, which makes it difficult to analyze the dissemination information of fake news. (4) In many cases, a user's decision to retweet a news is influenced not only by the previous one but also by all the users who have retweeted the news.

Facing these challenges, many scholars have done a lot of research. To address the misleading and generalized characteristics of fake news (difficulties (1) and (2)), scholars extract relevant information as subinformation based on the social context of news, such as dissemination paths [2, 3] and information of users who retweeted the news [4, 5], to assist the model in detecting news authenticity. Other scholars rely on machine learning to analyze the language style of news texts [6, 7] to find the commonality in the expression style of fake news. To address the complexity of news dissemination (difficulty (3)), some scholars represent the dissemination paths in the form of graphs and use graph neural networks to analyze [8]. Other scholars have analyzed the time series of news dissemination and mined temporal information [9]. However, most of methods extract single-modal features from the content or social context of news or simply combine textual features with social context features to obtain multimodal features. They all fail to realize that there is a large amount of contextual knowledge hidden in news, and the false knowledge conveyed by fake news is contrary to the existing knowledge system, ignoring the help brought by external knowledge system. In addition, when some models analyze the interaction relationships between users involved in retweeting the same news, they simply assume that users are only connected to the previous one in the propagation chain. In fact, users do not only see the comments of the previous one when reading a news but they can see the comments of multiple users, so their retweeting behavior is influenced by multiple users in the communication chain.

To address these problems, this paper applies the combination of external knowledge and GCN to the detection of fake news. The knowledge concepts in the news are extracted and linked to the wiki knowledge graph, and the corresponding knowledge entities and entity contexts in the knowledge graph are identified and introduced into the model as external knowledge. Meanwhile, this paper is inspired by the "similarity effect" in social psychology, and the GCN is weighted according to the similarity of user features to simulate more realistic user interaction relationships. Finally, a multimodal fake news detection model, A-KWGCN, based on external knowledge and weighted GCN is constructed by fusing multimodal features based on coattention.

The contributions of this paper are as follows: (1) while analyzing news texts, this paper introduces a variety of auxiliary information (such as external knowledge, user interactions, and communication chain), which jointly help the model detect fake news from multiple perspectives. (2) This paper introduces the concept of "similarity effect" in social psychology to weight the GCN, and making it more consistent with the realistic situation.

The rest of this paper is arranged as follows. The related research is presented in Section 2. The A-KWGCN model is described in Section 3. The experimental results and analysis

are described in Section 4. Finally, Section 5 introduces conclusions and future work.

2. Related Work

Early fake news detection mainly relied on manual detection, and websites such as Snopes and PolitiFact [10] employed multiple experts to check the authenticity of political news. Although the manual detection method can accurately identify fake news, it is labor-intensive and time-consuming, and the number of news that can be detected is very limited. The manual detection approach can no longer meet the needs of the rapidly expanding number of news with the emergence of the "Internet news era." The advancement of artificial intelligence technology has opened up new possibilities for detecting fake news, and computers' powerful computational capacity can swiftly and intelligently evaluate the authenticity of news. This kind of research has become one of the hot research works in academia. The starting points for solving the problem of fake news detection through artificial intelligence techniques are generally divided into two categories: fake news text-based detection and social media-based detection [11].

2.1. Text-Based Fake News Detection. Text-based methods for detecting fake news mainly rely on analyzing the content of news. Przybyla [7] created a style-based classifier for topic extraction and style analysis of news texts and compared it with a CNN-based classifier to demonstrate the former's effectiveness in capturing emotional elements in news texts. Based on the experimental results, the authors of the paper also proposed the idea that fake news tends to use some dramatic, uncritical, and lurid words for news reporting. Karimi and Tang [12] proposed a model based on hierarchical discourse-level structure for fake news detection, which learns and constructs real and fake news in an automated, data-driven way. Zhang and Li [13] built a cross-modal semantic deviation detection model based on the differences between the text and image of fake news and used semantic distance to detect fake news. Nasir et al. [14] proposed a hybrid CNN-RNN model by using a CNN layer of Conv1D for extracting the local features that reside at the text level and using a RNN layer for learning the long-term dependencies of the local features. Kaliyar et al. [15] proposed a BERT-based detection approach (FakeBERT) by combining different parallel blocks of the single-layer CNNs with the BERT, which improved the performance of fake news detection with the powerful ability to capture semantic and long-distance dependencies in sentences.

2.2. Social Context-Based Fake News Detection. Social media-based fake news detection methods mainly use the social context formed by the spread of fake news on social media. Liang et al. [2] focused on the influence of user features on the fake news detection model. By extracting 5 new user features and combining them with verified effective features of existing users, they constructed a brand new user feature set and applied it to the authenticity detection of weibo. Jin

et al. [3] proposed a fake news detection model integrating multimodal features, using different feature extraction methods and machine learning detection models for different types of features. Bian et al. [8] first applied GCN to the task of fake news detection. To study the propagation structure and dispersion structure of fake news, the authors used top-down and bottom-up GCNs to construct a BiGCN detection model and used DropEdge to mitigate the overfitting problem. Chen and Freire [16] focused on the detection of fake news domains (fake news domains refer to original websites that publish fake news), mined the sharing structure of Twitter users to find politically related websites, and constructed domain interaction maps by using real-time social networks. The topic-Agnostic classifier was used to score, rank, and classify the sites found, and finally, an active detection model of fake news domain was constructed. Raza and Ding [17] started from users' behaviors and calculated the aggregate responses of all users on each news article to understand a user's view on a news article and detect fake news by analyzing users' interactions with news. Davoudi et al. [18] proposed a new model to determine news veracity based on the analysis of the propagation tree and stance network features dynamically. Kaliyar et al. [19] combined news, user, and community information to form a tensor representing social context. The news content was fused with the tensor, and the coupling matrix tensor factorization was used to obtain the representation of news content and social context.

However, in terms of news text, most existing models have not sufficiently mined the news text information. These models usually focus on the mining of semantic features, ignoring the background knowledge hidden in the news content and intrinsic relationship with the external knowledge. In terms of communication chain, existing models tend to emphasize the single chain from the previous communication node, ignoring the intricate interaction and mutual influence among users in reality. Therefore, this paper introduces external knowledge features based on wiki knowledge graph and extracts user interaction features by building a weighted graph convolutional network based on "similarity effect," and constructs a detect model, A-KWGCN, that incorporates multimodal information to detect fake news.

3. A-KWGCN Model

In order to solve the above problems, this paper constructs an attention-based know-aware weighted graph convolutional neural (A-KWGCN) to detect fake news. The model introduces knowledge graph as auxiliary information through knowledge mining and entity linking techniques to help the model analyze the textual content of news. GCN is used to analyze the relationship network between users who are in the propagation chain. This paper fully analyzed the mutual influence relationship between users in the process of social network information dissemination. Based on the theory of "similarity effect" in psychology, the interaction between different nodes in the communication chain is reflected by a weighting algorithm, and the Laplace matrix in

GCN is improved accordingly, so that the model can more realistically reflect the relationship and influence between users in real information dissemination. Meanwhile, BERT is used to extract text features and GRU is used to extract propagation features. For the above multiple information, this paper uses the coattention mechanism for relevance analysis and finally splices the features and inputs them into the classifier for probability prediction. The specific framework of the model is shown in Figure 1.

The implementation of this framework mainly includes the following steps: ① data collection and preprocessing, ② knowledge mining and feature extraction of source information text, ③ text feature and propagation feature extraction based on BERT and GRU, ④ user interaction feature extraction based on weighted GCN, ⑤ correlation analysis of extracted features based on coattention, ⑥ fake news detection and classification.

3.1. Text Preprocessing and User Data Collection. News source text often contains a lot of redundant information, such as HTML codes, URLs, special symbols, and stop words. In the experiment, these redundant information and noise must be removed to retain important information, so as to ensure that useless information will not affect the classification accuracy of the model. This process is called text cleansing. For the cleaned text, the third party library NLTK is used for word segmentation. After the above processing, the news text data are represented as $S = \{s_1, s_2, s_3, \dots, s_n\}$, where $s_i (i = 1, 2, 3, \dots, n)$ represents the i -th preprocessed news text.

User information is one of the important information to achieve intelligent detection of fake news. For the news text collection S , the users who have forwarded these news form the user collection $U = \{u_1, u_2, u_3, \dots, u_m\}$, and the personal information of each user is collected to form the user feature matrix $X_u = \{X_{u1}, X_{u2}, X_{u3}, \dots, X_{um}\}$, where X_{uj} corresponds to u_j , u_j denotes the j -th user, and X_{uj} denotes the feature matrix corresponding to user u_j . The user feature matrix X_{uj} specifically contains the following information: ① the number of words of the user's nickname, ② the number of words of the user's profile, ③ the number of the user's followers, ④ the number of the user's followers, ⑤ the number of tweets posted by the user, ⑥ whether the user is allowed to share his geolocation, ⑦ whether the user is verified by his real name, ⑧ the number of years the user's account has been used, ⑨ the time elapsed from when the news was published to when the user retweeted the news, ⑩ whether the user account is protected or not. The final result is $X_{uj} \in R^v$, where v represents the number of features, and $v = 10$ in this paper.

3.2. Knowledge Mining and Feature Extraction of Source Information Text. Based on the preprocessed news text data obtained in Section 3.1, external knowledge is introduced as auxiliary information for fake news detection. Since fake news often spreads misleading and erroneous information, which is likely to contradict the existing human knowledge system, the introduction of external knowledge system is of

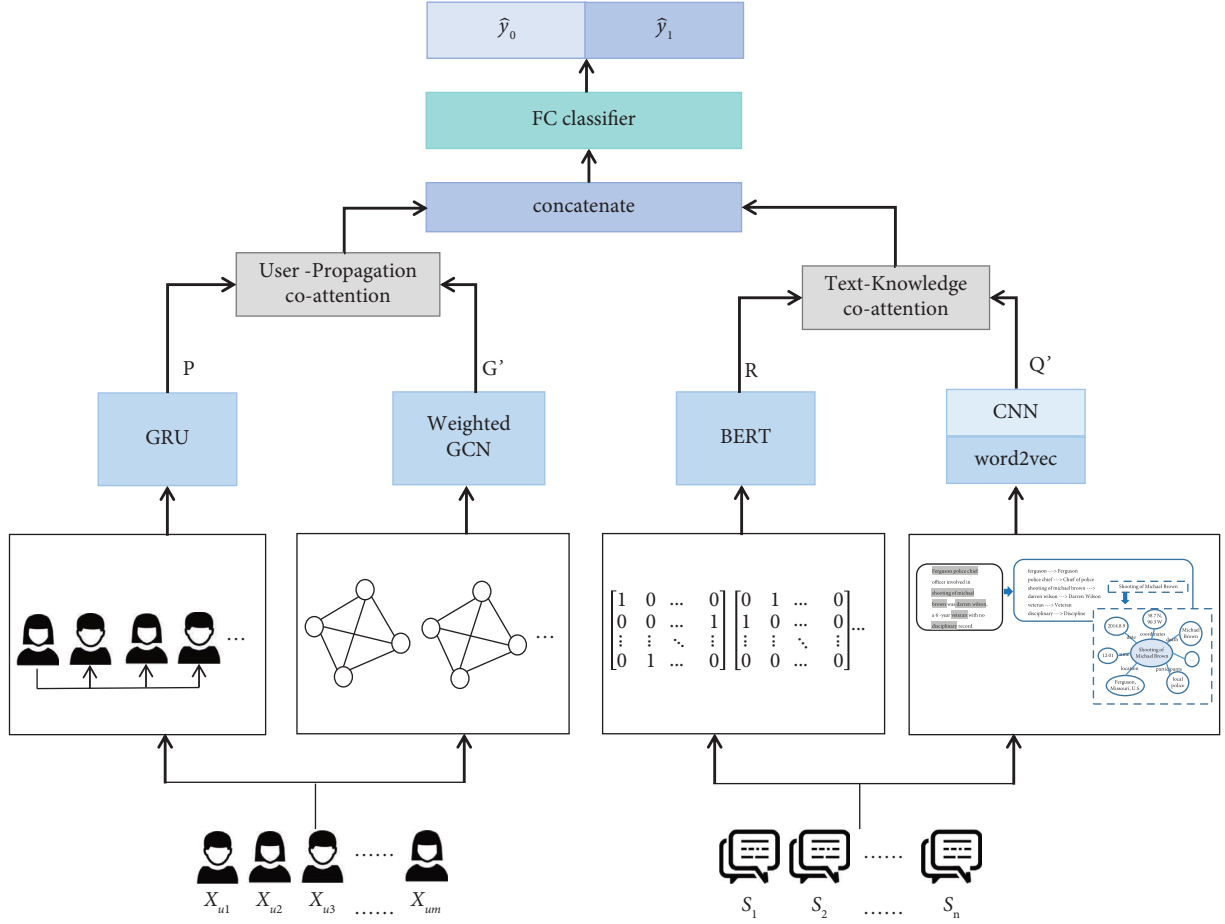


FIGURE 1: Structural framework of A-KWGCN model.

positive significance for detecting the truthfulness of news. In order to achieve the introduction of external knowledge, the following operations are required: textual knowledge extraction, knowledge entity linking, and knowledge feature extraction.

3.2.1. Textual Knowledge Extraction. Knowledge extraction is the process of extracting the knowledge embedded in the source information through identification, analysis, and induction. The purpose of this step is to mine words from news texts that may become knowledge mentions for subsequent linkage with entities in the knowledge graph. For the preprocessed news text data in Section 3.1, the third-party library, tagme, is used to extract the knowledge mentions from the text. Taking a specific tweet as an example, its final extracted knowledge mentions are shown in Figure 2.

3.2.2. Knowledge Entity Linking. Knowledge entity linking is the process of linking the knowledge mentions extracted from the text with the knowledge entities in the knowledge graph. The entity linking work needs to overcome two difficulties: (i) a mention corresponds to different knowledge entities, such as apple can link fruit apple or mobile phone brand apple. (ii) A knowledge entity corresponds to different

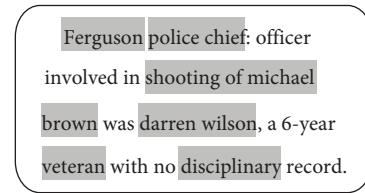


FIGURE 2: Knowledge extraction results display.

mentions, for example, Beijing and Peking can be linked with the knowledge entity Beijing. To address the above difficulties, this paper adopts candidate entity generation (CEG) and entity disambiguation (ED) methods mentioned by Shen et al. [20]. The sequence of knowledge entities $E = \{e_1, e_2, e_3, \dots, e_n\}$ for each news article is obtained, where n denotes the number of entities contained in each news. Meanwhile, in order to obtain richer external knowledge, for each knowledge entity e_i , this paper selects 10 neighboring entities that are closely connected to it to form the entity context:

$$\text{en}(e_i) = \{\text{en} | (\text{en} \sim e_i) \in \text{KG}\}, \quad (1)$$

where $\text{en} \sim e_i$ denotes that there exists a mutual relationship between entity en and entity e_i , KG denotes the knowledge graph, and $\text{en}(e_i)$ denotes the entity context of knowledge

entity e_i . For the sequence of knowledge entities $E = \{e_1, e_2, e_3, \dots, e_n\}$, after obtaining the entity context of each entity, the entity context sequence $EN = \{en(e_1), en(e_2), en(e_3), \dots, en(e_n)\}$ can be formed. Taking the news shown in Figure 2 as an example, its entity link results are shown in Figure 3.

3.2.3. Knowledge Feature Extraction. Knowledge feature extraction refers to the process of converting textual knowledge entities, which cannot be recognized by computers, into numerical features that can be analyzed and computed. For the knowledge entity sequence E and entity context sequence EN obtained from the “knowledge entity linking,” the computer cannot directly compute and analyze E and EN and needs to use word2vec technique to transform them into entity embedding $E' = \{e'_1, e'_2, e'_3, \dots, e'_n\}$ and entity context embedding $EN' = \{en'(e_1), en'(e_2), en'(e_3), \dots, en'(e_n)\}$, and $E', EN' \in R^{n \times d}$, where n is the number of entities, and d is the embedding dimension. Since the entity context $en(e_i)$ is composed of multiple entities en , computing $en'(e_i)$ is a process of mean calculation:

$$en'(e_i) = \frac{1}{\text{num}(en(e_i))} \sum_{j=0}^{\text{num}(en(e_i))} en'_j, \quad (2)$$

where $en'(e_i)$ denotes the entity context embedding vector of entity e_i , $\text{num}(en(e_i))$ denotes the number of neighboring entities contained in $en(e_i)$ (in this paper, $\text{num}(en(e_i)) = 10$), and en'_j denotes the embedding vector corresponding to the j -th neighboring entity in $en(e_i)$. Subsequently, in order to obtain the knowledge embedding EE' that contains both knowledge entities and entity contexts, it is necessary to fuse E' and EN' . For the sake of computational simplicity, this paper fuses E' and EN' by computing the average value to obtain the sequence of knowledge embedding $EE' = \{ee'_1, ee'_2, ee'_3, \dots, ee'_n\}$, which is calculated as shown in the following equation:

$$ee'_i = \frac{1}{2} (e'_i + en'(e_i)). \quad (3)$$

Finally, the obtained knowledge embedding sequence EE' is input into the CNN, and the knowledge feature matrix Q' , which contains both knowledge entity and entity context information, is output and used in the subsequent classification calculation.

3.3. Text Feature and Propagation Feature Extraction Based on BERT and GRU. In order to be able to obtain information about the news content for fake news detection, features of the news text need to be extracted. The preprocessed news text data s_i obtained in Section 3.1 is one-hot coded to obtain a high-dimensional 0-1 matrix. Since the length of each text is different from each other, the maximum text length L needs to be set to make the 0-1 matrix dimension of each news uniform, and the part smaller than the L is filled with 0 to get $OH = \{oh_1, oh_2, \dots, oh_L\}$, where oh_i denotes the one-hot encoding of the i -th word. The OH is encoded by

using word2vec to obtain the word embedding, where V denotes the word embedding $V = \{v_1, v_2, \dots, v_L\}$:

$$V = \text{word2vec}(OH). \quad (4)$$

BERT is a deep bidirectional language representation model, which has great advantages in obtaining the deep semantic information of text [15, 21–23]. As shown in equation (5), V is input into BERT to obtain the textual features of news $R = \{r_1, r_2, \dots, r_L\}$, $R \in R^{d \times L}$, where d denotes the dimensionality.

$$r_i = \text{BERT}(v_i). \quad (5)$$

For the user information X obtained in Section 3.1, the propagation features $P = \{p_1, p_2, \dots, p_\eta\}$ of the news can be obtained using the GRU model, with $p_i = \text{GRU}(X_{ui})$. Note that, for each of the news, the number of retweet users should be the same, and the specific number is set to η . The order of X_{ui} should be the same as the sequence of retweets.

3.4. User Interaction Feature Extraction Based on Weighted GCN. The purpose of crawling user information in Section 3.1 is to obtain information about user interactions. In this paper, we hypothesize that the interaction information between users helps in fake news detection. This hypothesis is validated in Section 4.4. To obtain potential interaction information among all users who retweeted news s_i , this paper uses GCN to extract user interaction features. In previous literature, only the interaction between a user and the previous one in the propagation chain is studied, but in practice, a user's decision to retweet a news item may not be influenced by the previous one only, but by all users who have retweeted, the news item before that. Take the social networking platform Twitter as an example, users can view the tweets of any other users and retweet the tweet they agree with. Therefore, in the real news dissemination process, it is very common that the target user interacts with multiple users, and we cannot ignore the positive impact of this phenomenon on fake news detection. Based on the above, a weighted GCN is created in this paper. First, a user social network graph $G = (V, \mathcal{E})$ is created, where V denotes the node, and \mathcal{E} denotes the edge. In this paper, node V is the set of users who have retweeted news s_i , and \mathcal{E} is the interaction relationship that exists between users. Since the interaction between users in the actual communication process is difficult to obtain, and we still want to simulate a more realistic user interaction situation, the “similarity effect” in social psychology is introduced. The concept suggests that humans are more inclined to accept the views of those who are similar to them [24]. Therefore, the user social network graph G created in this paper is a fully connected graph, where each edge \mathcal{E}_{ij} is associated with a weight w_{ij} . The specific value of the weight w_{ij} is obtained by calculating the cos similarity between the feature vectors X_{ui} , X_{uj} (obtained from Section 3.1) of user i and user j . The calculation formula is shown in equation (6). The final weight matrix $W_{\mathcal{E}} = [w_{ij}] \in R^{\eta \times \eta}$ is obtained, where η denotes the number of users.

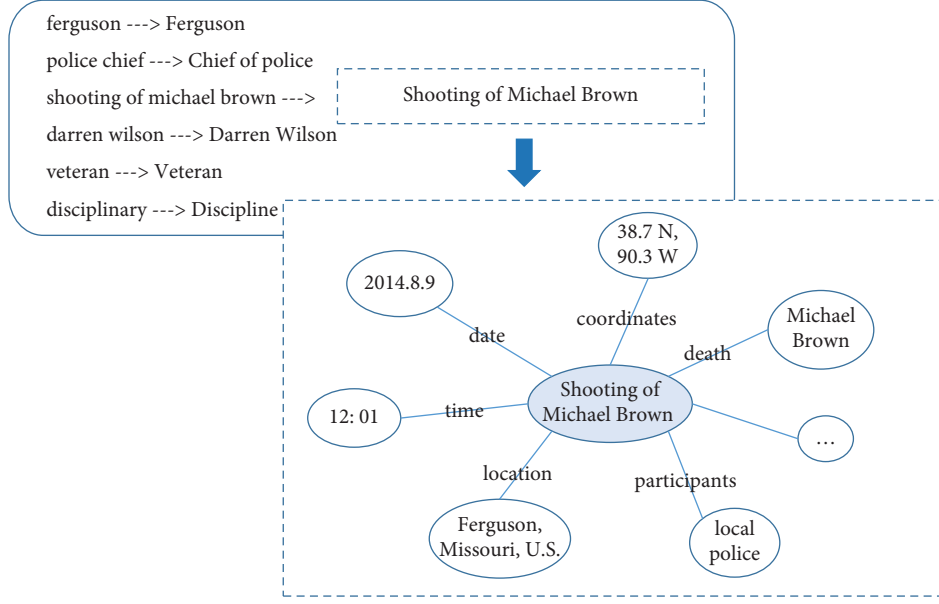


FIGURE 3: Entity linking result display.

$$w_{ij} = \frac{X_{ui} \cdot X_{uj}}{\|X_{ui}\| \|X_{uj}\|}. \quad (6)$$

GCN is a multilayer neural network that processes non-Euclidean data and can generate embedding vectors of nodes by obtaining information from their direct and indirect neighbors through the operation of stacked hierarchical convolution. In this paper, we use GCN to analyze the created user social network graph G . In the GCN network, the propagation formula between layers is shown in the following equation:

$$H^{(l+1)} = \sigma(D^{-(1/2)} A D^{-(1/2)} H^{(l)} W^{(l)}), \quad (7)$$

where $D^{-(1/2)} A D^{-(1/2)}$ is the Laplace matrix, A is the dependency matrix, D is the degree matrix of A , $H^{(l)}$ is the feature vector of the l -th layer, $H^{(0)}$ is set to the user feature matrix X_u obtained in Section 3.1, and $W^{(l)}$ is the weight matrix of each layer. σ is a nonlinear activation function, such as ReLU function. By the above operation, the user interaction features $G' = \{g_1, g_2, \dots, g_n\}$ can be obtained.

3.5. Correlation Analysis of Extracted Features Based on Coattention. For the knowledge features (Section 3.2), text features (Section 3.3), user interaction features (Section 3.4), and propagation feature (Section 3.3), this paper uses the method of Hu et al. [10] to construct coattention mechanism to learn the correlation between source tweet text information and knowledge information, and the joint influence between propagation feature and user interaction information, respectively. Take the analysis of source tweet text information and external knowledge information as an example. First, the proximity matrix F is calculated as shown in the following equation:

$$F = \tanh(RW_f Q'), \quad (8)$$

where R is the text feature obtained in Section 3.3, Q' is the external knowledge feature obtained in subsection 3.2, and W_f is the weight matrix.

Then, the proximity matrix F is brought into equation (9) to calculate and get the attention maps $H^R, H^{Q'}$ for the source tweet text information and external knowledge information, respectively.

$$\begin{aligned} H^R &= \tanh(W_R R + (W_{Q'} Q') F^T), \\ H^{Q'} &= \tanh(W_{Q'} Q' + (W_R R) F), \end{aligned} \quad (9)$$

where $W_R, W_{Q'}$ are the weight matrices.

The obtained $H^R, H^{Q'}$ are brought into equation (10) for calculation to obtain the attention probabilities a^R and $a^{Q'}$ of each word and knowledge. The weighted summation obtains the attention vectors \hat{R} and $\hat{Q'}$ for text information and external knowledge information, as shown in the following equation (11). In the same way, the attention vectors \hat{P} and $\hat{G'}$ of the propagation information and user interaction information can be obtained.

$$\begin{aligned} a^R &= \text{softmax}(w_{hr}^T H^R), \\ a^{Q'} &= \text{softmax}(w_{hQ'}^T H^{Q'}), \end{aligned} \quad (10)$$

$$\begin{aligned} \hat{R}_1 &= \sum_{i=1}^L a_i^R r_i, \\ \hat{Q'} &= \sum_{j=1}^n a_j^{Q'} q_j, \end{aligned} \quad (11)$$

where $w_{hr}, w_{hQ'}$ are the weight matrix, and r_i, q_j are the learned coattention feature vectors.

3.6. Fake News Detection and Classification. The \hat{P} , \hat{G}' , \hat{R} , and \hat{Q} obtained in 3.5 are spliced and input to a classifier consisting of a fully connected neural network to obtain the final classification probability distribution $y = [\hat{y}_0, \hat{y}_1]$, where \hat{y}_0 denotes the probability that the target news is classified as true news y_0 and \hat{y}_1 denotes the probability that the target news is classified as fake news y_1 . And the loss function formula is used as shown in the following equation:

$$\Gamma = -y \log(\hat{y}_1) - (1 - y) \log(1 - \hat{y}_0). \quad (12)$$

4. Experimental Results and Analysis

The codes of this experiment were written using Python 3.7 and run through Google artificial intelligence platform.

4.1. Dataset. In this paper, two datasets, Twitter15 and Twitter16 [25], are selected as the original datasets. The datasets contain fact-based source tweet text, labels, and user retweet sequences. Personal information is not given in the dataset due to Twitter’s protection of users’ personal information, and thus, the experimenter needs to crawl it by using the Twitter API according to the user ids in the dataset. The user information to be crawled in this paper is described in Section 3.1. The data with “true” and “fake” labels are selected for analysis, and the details are shown in Table 1. The dataset is divided into a training set and a test set in the ratio of 7 : 3.

4.2. Parameter Setting. In this paper, we detect the authenticity of fake news and classify them into two categories: “true” and “fake.” The parameter settings of the model in this experiment are shown in Table 2. Since the length of each tweet text is different and the number of retweeted users per text is also different, the maximum text length is set to 30, and the number of users per text is set to 40. For texts with more than 40 retweeted users, only the first 40 users are retained; for texts with less than 40 retweeted users, a random sample of existing users is required to fill the gap until the number of users reaches 40. Kipf and Welling [26] pointed out that the number of layers of GCN should be controlled in 1~3. In this paper, the number of layers of GCN is set to 2. The softmax function is chosen as the activation function of the fully connected neural network classifier, and the ReLU function is used as the activation function of GCN, which is σ in equation (7). Meanwhile, this paper introduces external knowledge based on the wiki knowledge graph. The wiki knowledge graph is free and open, and its contents are all structured information that can provide support for A-KWGCN model.

4.3. Comparison of Different Models. In this paper, the experimental results of A-KWGCN are compared with some current state-of-the-art models and some representative baseline methods to demonstrate the performance of the model proposed in this paper. The comparison models chosen in this paper are shown as follows:

TABLE 1: Structure of the dataset.

	Twitter15	Twitter16
Number of source tweets	742	412
Labeling is true	372	205
Labeling is fake	370	207
Number of users	190,868	115,036

DTC [27]: A decision tree based detection model, which achieves automatic news trustworthiness assessment by tagging the topics of news content and extracting relevant topic-word features.

SVM-TS [28]: A detection model built on support vector machine, which captures time series features and combines them with source tweets to jointly analyze the truthfulness of news.

mGRU [29]: A detection model built on an improved recurrent neural network is used to capture the contextual information of tweets over time for learning temporal features and textual features.

CSI [30]: A multimodal fake news detection model that combines and analyzes three types of information: source text information, user information, and information about the group of users involved in retweeting news.

DEFEND [31]: An advanced fake news detection model that analyzes source news and user profiles. And this model is dedicated to learning the relationship between news data and user data through coattention mechanism.

GCAN [9]: A graph-aware network based model. It predicts the authenticity of the target news by analyzing the source text and the corresponding sequence of retweeted users. It is a relatively advanced model for fake news detection.

Rumor2vec [32]: A model detects fake news based on the text content and the propagation structure. It presents the concept of the union graph to incorporate propagation structures of all tweets and leverages network embedding to learn representations of nodes in the union graph.

GAN [33]: A model uses graph attention neural network model to learn text features and syntactic relations to detect fake news.

In this paper, F1-score, Recall, Precision, and Accuracy are chosen as evaluation metrics to compare the performance between different models. The conducted train test is repeated 10 times, with the average values provided. And the specific results of the comparison experiments are shown in Table 3. It is shown that A-KWGCN outperforms other models. Its performance is improved by 5.3% and 9.4% on average on the Twitter15 and Twitter16 datasets, respectively. And it demonstrates the effectiveness of the A-KWGCN model in detecting fake news.

We can also find the following conclusions in the table:

- (1) The models using deep learning methods generally outperform those using hand-crafted features (e.g., DTC, SVM-TS). This is because deep learning based

TABLE 2: Parameter settings.

Stage	Parameter	Value
Preprocessing	Number of users per tweet	40
	Maximum text length	30
	Dimension of knowledge embedding	300
Feature acquisition	CNN output dimension	32
	Filter size	3
	BERT output dimension	32
	GRU output dimension	32
	GCN output dimension	32
	Number of GCN layers	2
Training	Coattention output dimension	64
	Optimizer	Adam
	Batch size	64
	Training cycle	100
	Learning rate	0.001

TABLE 3: Experimental results.

Model	Accuracy	F1-score	Precision	Recall
(a) Experimental results on Twitter 15				
DTC	0.495	0.495	0.496	0.481
SVM-TS	0.520	0.519	0.520	0.519
mGRU	0.555	0.510	0.515	0.515
CSI	0.699	0.717	0.699	0.687
dEFEND	0.738	0.654	0.658	0.661
Rumor2vec	0.791	0.782	0.780	0.785
GAN	0.856	0.868	0.847	0.853
GCAN	0.877	0.825	0.826	0.830
A-KWGCN	0.905	0.891	0.895	0.888
(b) Experimental results on Twitter 16				
DTC	0.561	0.562	0.575	0.537
SVM-TS	0.693	0.692	0.693	0.691
mGRU	0.661	0.556	0.560	0.562
CSI	0.661	0.630	0.632	0.631
dEFEND	0.702	0.631	0.637	0.638
Rumor2vec	0.848	0.840	0.827	0.836
GAN	0.854	0.870	0.861	0.847
GCAN	0.908	0.759	0.759	0.763
A-KWGCN	0.930	0.879	0.882	0.879

models are able to extract higher level news text features and user features, which also proves that introducing deep learning methods for detecting fake news is an effective research direction.

- (2) The CSI, dEFEND, Rumor2vec, and GAN combine information from multiple modalities (e.g., text features, user features, and retweet sequence features), and its performance is significantly better than models that analyze single modal feature (e.g., DTC, SVM-TS, mGRU), which also proves that the focus of fake news detection is on the selection and analysis of multi-modal features. Building richer multimodal information features is an effective way to improve model detection performance.
- (3) GCAN analyzes text, user, and forwarding sequence features just like CSI. However, compared to CSI, which simplifies user relationships to Euclidean data, GCAN constructs a user social network and

introduces graph neural networks into the model to create a graph-aware fake news detection model, and its performance is also greatly improved compared to CSI. It also proves that the simplification of complex data will affect the detection effect of the model to some extent, and the improvement of feature extraction means can bring help to detect fake news. Among all the models, the A-KWGCN has the highest classification accuracy, which also proves that mining the background knowledge of news, introducing external knowledge features as auxiliary information and improving the GCN to mine the user interaction features that are closer to reality can bring positive effects on the detection of fake news, which also provides new research ideas and models for the detection and classification of fake news.

4.4. Ablation Experiment. In order to verify the effectiveness of each module in this model for detecting news authenticity, this paper conducts ablation experiments based on the original model. Three versions of “A-KWGCN/G,” “A-KWGCN/K,” “A-KWGCN/UI,” and “A-KWGCN/ALL” were created based on A-KWGCN. The details are explained as follows:

A-KWGCN/G: Remove the graph-aware module of the original model. Replace the weighted GCN with CNN, and instead of creating a graph of user social relations, the CNN is used to convolve and analyze the user data obtained in subsection 3.1 directly. The other parts of the model remain unchanged.

A-KWGCN/UI: The user-interaction module of the original model is deleted. With the deletion of user interaction information, the user-propagation coattention mechanism also needs to be deleted.

A-KWGCN/K: The knowledge-aware module of the original model is deleted. With the deletion of external knowledge information, the text-knowledge coattention mechanism also needs to be deleted.

A-KWGCN/ALL: By removing all auxiliary information including external knowledge, user interactions, and propagation information, the model analyzes only the textual content of the news.

In the “A-KWGCN/G” model, the filter size of CNN is set to 3, and the output dimension is set to 32. “A-KWGCN/G,” “A-KWGCN/K,” “A-KWGCN/UI,” and “A-KWGCN/ALL” are used to verify the role of weighted GCN in extracting user interaction information, the role of introducing external knowledge system, the role of analyzing user interactions, and the role of news content, respectively. The results of the ablation experiments are displayed in Figure 4. It can be observed that the model accuracy is in the order of “A-KWGCN” > “A-KWGCN/K” > “A-KWGCN/G” > “A-KWGCN/UI” > “A-KWGCN/ALL.” This illustrates that the graph-aware module, the knowledge-aware module, the user-interaction module, and the textual content of the news all provide great help to the detection of fake news.

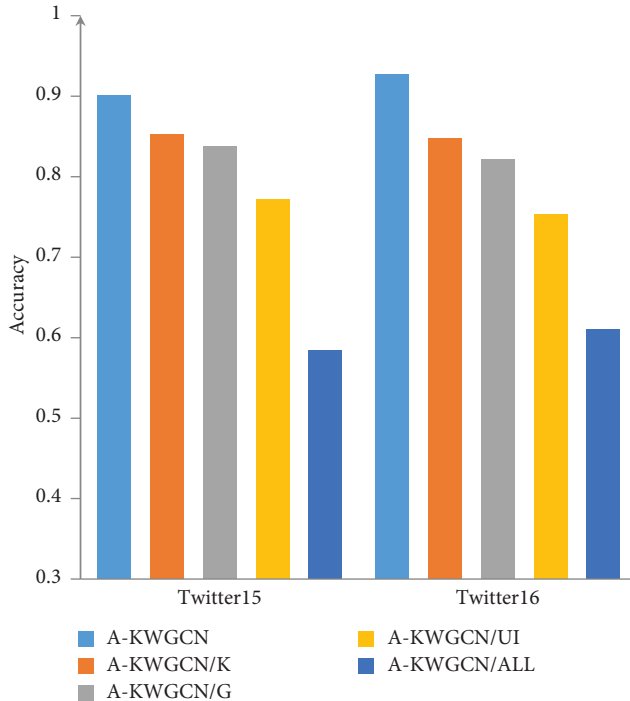


FIGURE 4: Results of ablation experiment.

Introducing external knowledge systems and extracting user interaction information that is closer to the real propagation process as auxiliary information are feasible methods. And the decrease in the accuracy of the A-KWGCN/UI model also proves the scientific validity of the hypothesis that “user interaction information is helpful for false news detection.” At the same time, the detection accuracy of “A-KWGCN/ALL” is the lowest, which indicates that there is a significant decrease in the classification accuracy of the model after removing all the auxiliary information. It also proves that the positive impact brought by the auxiliary information on the model to detect fake news cannot be ignored, and this also provides new ideas for the construction of future fake news detection models.

5. Conclusion and Future Work

While social media provides the convenience for news dissemination, it also leads to the rapid spread of fake news. The intelligent detection of fake news is helpful to develop a healthy network environment and keep society stable. At present, the problems in the fake news detection models are as follows: (1) the mining and refining of fake news content are not sufficient, and the background knowledge hidden in the news content and its internal relationship with the existing knowledge system are often ignored. (2) In the analysis of the user propagation relationships, existing models simply assume that users only interact to the previous one, ignoring the intricate interaction relationship among users. To address these problems and achieve intelligent detection of fake news, this paper proposes a multimodal fake news detection model by mining the hidden knowledge concepts in the news content, and linking

them to wiki knowledge graph to introduce knowledge entities and entity contexts as auxiliary information to help the model detect fake news. Meanwhile, this paper improves the GCN based on the “similarity effect” in social psychology. And a weighted GCN is constructed by calculating user feature similarity to simulate the real interaction among users. In this paper, two public datasets, Twitter15 and Twitter16, are selected for comparison experiments and discussions, and the model is compared with several models in terms of classification accuracy and efficiency. According to the experiment results, the classification accuracy of this model reaches 0.905 and 0.930 in the two datasets, respectively, and the model performs better than other models in four evaluation metrics and obtains the highest classification accuracy. Also, ablation experiments and analytical discussions are conducted and demonstrated the significant role of knowledge graph module and the weighted GCN module in classification efficiency. The research in this paper provides new research ideas and models for the detection and classification of fake information.

The ability of the model to achieve better classification accuracy is important, but we argue that it is equally important to explain the classification results, i.e., why the target news is judged as true news or fake news. However, it is missing in this experiment. In future work, we will work on achieving interpretable detection of fake news by analyzing the reasons why the target news is judged as fake (or true) in terms of news content, user characteristics, and external knowledge and by transforming complex calculation into information that people can understand, and improving people’s trust in the classification results.

Data Availability

The Twitter15, 16 data used to support the findings of this study have been deposited in the <https://www.dropbox.com/s/7ewzdrbelpmrnxu/rumdetect2017.zip?dl=0>.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research was funded by the General Project of Philosophy and Social Science of Heilongjiang Province, grant number: 20SHB080.

References

- [1] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” *The Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211–236, 2017.
- [2] G. Liang, W. He, C. Xu, L. Chen, and J. Zeng, “Rumor identification in microblogging systems based on users’ behavior,” *IEEE Transactions on Computational Social Systems*, vol. 2, no. 3, pp. 99–108, 2015.
- [3] Z. Jin, J. Cao, B. Wang, R. Wang, and Y. D. Zhang, “Social multimedia rumor detection based on multi-modal features,” *Journal of Nanjing University of Information Science & Technology (Natural Science)*, vol. 6, pp. 583–592, 2017.

- [4] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *Proceedings of the 2018 IEEE conference on multimedia information processing and retrieval (MIPR)*, pp. 430–435, IEEE, Miami, FL, USA, April, 2018.
- [5] Y. Dou, K. Shu, C. Xia, P. S. Yu, and L. Sun, "User preference-aware fake news detection," 2021, <https://arxiv.org/abs/2104.12259>.
- [6] K. Popat, "Assessing the credibility of claims on the web," in *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 735–739, Geneva, Switzerland, April, 2017.
- [7] P. Przybyla, "Capturing the style of fake news," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 1, pp. 490–497, New York, NY, USA, April, 2020.
- [8] T. Bian, X. Xiao, T. Xu et al., "Rumor detection on social media with bi-directional graph convolutional networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 1, pp. 549–556, New York, NY, USA, April, 2020.
- [9] Y. J. Lu and C. T. Li, "GCAN: graph-aware Co-attention networks for explainable fake news detection on social media," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 505–514, Seattle, WA, USA, July, 2020.
- [10] Y. Hu, "Effective fraud, weak correction? Preliminary study on the third party fact checking mechanism," *Dissemination of research and practice*, vol. 8, no. 2, pp. 43–73, 2018.
- [11] P. F. Liu, "Research and application of false News detection algorithm based on multi-modal feature and semantic enhancement," Master's thesis, Shandong University of Science and Technology, Shan Dong Sheng, China, 2020.
- [12] H. Karimi and J. Tang, "Learning hierarchical discourse-level structure for fake news detection," 2019, <https://arxiv.org/abs/1903.07389>.
- [13] G. B. Zhang and J. Li, "Social Media false news detection based on multi-modal content semantic consistency," *Data Analysis and Knowledge Discovery*, vol. 5, no. 5, pp. 21–29, 2021.
- [14] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: a hybrid CNN-RNN based deep learning approach," *International Journal of Information Management Data Insights*, vol. 1, no. 1, Article ID 100007, 2021.
- [15] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11765–11788, 2021.
- [16] Z. Chen and J. Freire, "Proactive discovery of fake news domains from real-time social media feeds," in *Proceedings of the 29th International World Wide Web Conference (WWW 2020)*, pp. 584–592, Association for Computing Machinery, New York, NY, USA, April, 2020.
- [17] S. Raza and C. Ding, "Fake news detection based on news content and social contexts: a transformer-based approach," *International Journal of Data Science and Analytics*, vol. 13, no. 4, pp. 335–362, 2022.
- [18] M. Davoudi, M. R. Moosavi, and M. H. Sadreddini, "DSS: a hybrid deep model for fake news detection using propagation tree and stance network," *Expert Systems with Applications*, vol. 198, Article ID 116635, 2022.
- [19] R. K. Kaliyar, A. Goswami, and P. Narang, "DeepFakE: improving fake news detection using tensor decomposition-based deep neural network," *The Journal of Supercomputing*, vol. 77, no. 2, pp. 1015–1037, 2021.
- [20] W. Shen, J. Wang, and J. Han, "Entity linking with a knowledge base: issues, techniques, and solutions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 2, pp. 443–460, 2015.
- [21] M. Szczepański, M. Pawlicki, R. Kozik, and M. Choraś, "New explainability method for BERT-based model in fake news detection," *Scientific Reports*, vol. 11, no. 1, Article ID 23705, 2021.
- [22] H. Jwa, D. Oh, K. Park, J. M. Kang, and H. Lim, "exbake: automatic fake news detection model based on bidirectional encoder representations from transformers (bert)," *Applied Sciences*, vol. 9, no. 19, p. 4062, 2019.
- [23] P. K. Verma, P. Agrawal, V. Madaan, and R. Prodan, "MCRED: multi-modal message credibility for fake news detection using BERT and CNN," *Journal of Ambient Intelligence and Humanized Computing*, vol. 27, pp. 1–13, 2022.
- [24] T. Zhu, "Psychology in interpersonal communication," 2009, <http://psychology.iresearchnet.com/industrial-organizational-psychology/group-dynamics/interpersonal-communication/>.
- [25] J. Ma, W. Gao, and K. F. Wong, "Detect rumors in microblog posts using propagation structure via kernel learning," in *Proceedings of the 55th annual meeting of the Association for Computational Linguistics (ACL 2017)*, Association for Computational Linguistics, Vancouver, Canada, June 2017.
- [26] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017, <https://arxiv.org/abs/1609.02907>.
- [27] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, pp. 675–684, New York, NY, USA, March, 2011.
- [28] J. Ma, W. Gao, Z. Wei, Y. Lu, and K. F. Wong, "Detect rumors using time series of social context information on microblogging websites," in *Proceedings of the 24th ACM international conference on information and knowledge management*, pp. 1751–1754, New York, NY, USA, October, 2015.
- [29] J. Ma, W. Gao, P. Mitra et al., "Detecting rumors from microblogs with recurrent neural networks," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016)*, New York, NY, USA, June 2016.
- [30] N. Ruchansky, S. Seo, and Y. Liu, "Csi: a hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM Conference on Information and Knowledge Management*, pp. 797–806, New York, NY, USA, November, 2017.
- [31] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "Defend: explainable fake news detection," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 395–405, New York, NY, USA, July, 2019.
- [32] K. Tu, C. Chen, C. Hou, J. Yuan, J. Li, and X. Yuan, "Rumor2vec: a rumor detection framework with joint text and propagation structure representation learning," *Information Sciences*, vol. 560, pp. 137–151, 2021.
- [33] Y. Lv, X. Sun, Y. Wen, and W. Wang, "Rumor detection based on graph attention network," in *ITM Web of Conferences*, vol. 47, p. 2033, EDP Sciences, Paris, France, 2022.