International Journal of Information Technologies and Systems Approach
Volume 17 • Issue 1 • January-December 2024

GMRD:
A Rumor Detection Model Based on Graph Convolutional Networks and Multimodal Features

Qian Li
Zhoukou Normal University, China

Laihang Yu
Zhoukou Normal University, China

Li Pan
UCSI University, Malaysia

ABSTRACT

The rapid development of social media has allowed people to access information through multiple channels, but social media has also become a breeding ground for rumors. Rumor detection models can effectively assess the credibility of information. However, current research mainly relies on text or combined text and image features, which may not be sufficient to capture complex feature information. Therefore, this paper proposes a rumor detection model based on the graph convolutional network (GCN) and multi-modal features. The proposed model constructs a knowledge graph (KG) and leverages the GCN to extract complex relationships between its nodes. Then, an interactive attention network is adopted to deeply integrate features. Furthermore, ResNet101 is utilized to extract salient features from images, addressing the challenges related to fully utilizing additional feature information and capturing text and image features at a deeper level to some extent. Multiple experiments conducted on datasets from Twitter and Weibo platforms demonstrate the efficacy of the proposed approach.

KEYWORDS

Graph Convolutional Network, Knowledge Graph, Rumor Detection

INTRODUCTION

Deliberately misleading information, also known as disinformation, has serious consequences for both society and individuals (Bin & Sun, 2022). Its dissemination on social media platforms is characterized by wider, faster, deeper, and broader reach. The immense negative consequences of disinformation have made it a pressing issue, gaining extensive attention from researchers. Therefore, studying the characteristics of false information dissemination on social networks and quickly identifying rumors hold significant implications for the development and governance of social media platforms (Lawson-Body et al., 2023).

The purpose of rumor detection is to effectively separate fact from fiction. The rumor detection methods include manual detection and automatic detection based on machine learning and deep learning. Manual identification relies on feedback from users and content authentication by reviewers, incurring substantial human and time costs. Given the vast amount of information in the era of big

DOI: 10.4018/IJITSA.348659

data, manual detection is prone to errors and omissions. Therefore, research has shifted towards automatic rumor detection methods. The first machine learning-based rumor detection model using a decision tree was proposed in 2011 to assess the credibility of rumors on Twitter[1]. While early machine learning-based rumor detection models achieved decent accuracy (ACC) in identifying rumors, they relied on time-consuming and laborious feature engineering. This approach struggles to meet the research needs of rumor detection in the big data era.

In recent years, the potent feature extraction capabilities of deep neural network (DNN) methods have compensated for the shortcomings of machine learning models. These methods have achieved fruitful research results in machine translation, sentiment analysis, and disease monitoring (Qin et al., 2022). Consequently, the construction of the DNN has gradually become the mainstream approach in rumor detection. The rise of neural language processing (NLP) has led to the development of attention mechanisms. Liao et al. (2018) proposed a method based on hierarchical attention networks, utilizing bidirectional GRUs with dual-layer attention mechanisms to automatically learn key information from text. Rumor researchers are increasingly recognizing the power of images to spread misinformation. Singhal et al. (2019) utilized VGG19 (Visual Geometry Group) and BERT (Bidirectional Encoder Representation from Transformer) for extracting visual and textual information, respectively. They then concatenated the textual and visual information for classification. However, the direct concatenation of visual and textual information is overly simplistic and struggles to fully leverage multimodal information.

To enhance the understanding of multimodal information in rumor detection models, several auxiliary tasks have been designed. K. Zhang et al. (2023) employed diverse event graphs as factual evidence and devised an effective strategy for generating different sub-graphs of the event graph. These sub-graphs were designed to naturally serve as counterfactual evidence, aiding in the detection of fake news. In recent years, several studies have incorporated external knowledge into rumor detection research to obtain comprehensive semantic representations. Dun et al. (2021) utilized named entity recognition for rumor detection, mapping the real-world entities mentioned in the text to entities within a knowledge graph (KG) to identify corresponding entities.

Current research on rumor detection relies on features such as images, text, and, in some cases, supplements from KGs. However, few studies simultaneously utilize all three features. Therefore, this paper proposes a novel method that innovatively incorporates image, text, and external knowledge in feature usage. The proposed framework utilizes external knowledge as a complement to textual features to improve the modeling effectiveness of textual features. Specifically, a rumor detection model based on graph convolutional networks (GCNs) and multimodal features is proposed. The proposed framework integrates external knowledge by extracting entity background knowledge and aggregating it with attention mechanisms to obtain semantically enriched external knowledge. GCNs and multi-head attention networks are employed to capture effective information. Additionally, an interaction attention network (INT-Att) is introduced to create a deep connection between this external knowledge and the original text features. ResNet101 is employed to effectively capture image information from pictures. Overall, this research makes the following contributions:

1. This paper proposes a rumor detection model based on graph convolutional networks and multimodal features GMRD, a novel rumor detection model that leverages GCNs and multimodal features. The proposed GMRD tackles key issues in rumor detection such as semantic understanding, multimodal information fusion, and model design.
2. By leveraging the ConceptNet KG, the proposed model accesses information from open-domain common sense knowledge, effectively enhancing the understanding of textual semantics. This enables more accurate identification of rumors within the text, aiding users in better discerning the authenticity of the information.
3. An interactive attention fusion network is introduced to facilitate interaction and fusion between textual and image features. This capability in integrating multimodal information enables the

model to comprehensively understand information within social media platforms and better identify rumors.

The remainder of this paper is organized as follows: The next section reviews the related research on rumor detection classification. The section after that describes the development of the GMRD model. The following section encompasses experiments and evaluations. The next section presents discussions, and the final section concludes the entire work.

RELATED WORKS

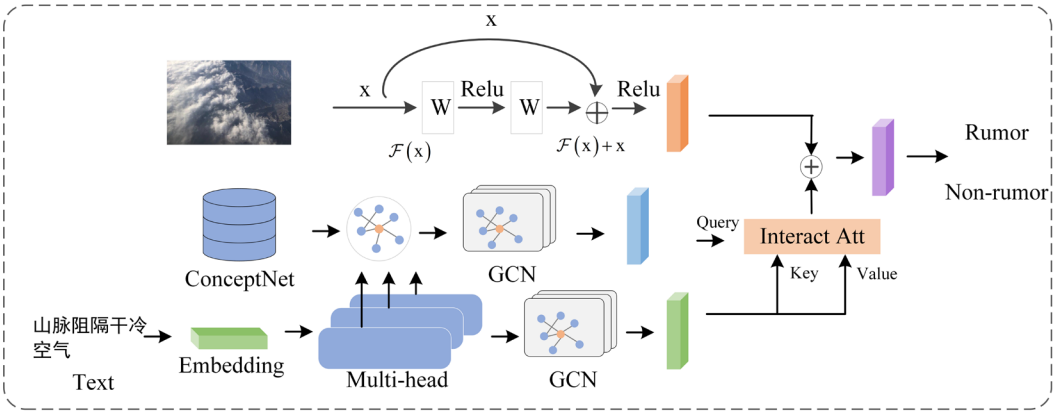
With the development of deep learning, scholars have gradually recognized its advantages over traditional machine learning. Deep learning possesses stronger representational capabilities, ACC, and applicability.

Ma et al. (2018) used a recurrent neural network (RNN) model for rumor detection on Weibo. They modeled text vectors with long short-term memory (LSTM) to effectively consider the context and contextual information of text. Liu et al. (2017) employed a character-level convolutional neural network (CNN) model, treating text as character sequences and mapping characters to vectors, which is followed by using convolutional and pooling layers to extract features. D. Lin et al. (2019) utilized LSTM to capture sequential contextual features of content for learning the falsehood of information and used CNN to learn the relationship. Zhou et al. (2018) employed CNN to automatically construct rumor features and used a gated recurrent unit (GRU) to explore information between Weibo posts. Moreover, GCN (Bai et al., 2021) and generative adversarial networks (Guo et al., 2021) have been effectively employed in the domain of rumor detection. In deep learning models, different types of information have different levels of importance. Therefore, researchers have introduced attention mechanisms into the problem of rumor detection. To detect highly attentive information, T. Chen et al. (2018) introduced attention mechanisms into RNNs to capture implicit and explicit characters from repetitive and variable Twitter information. Peng and Wang (2021) explored the temporal sequence background and sentiment polarity features of rumor lifecycles, utilizing a CNN model with spatial attention mechanisms for rumor detection and classification.

However, the use of images to propagate rumors has become increasingly prevalent, making it difficult to effectively identify rumor information solely through textual features. To tackle this issue, Weibo rumor detection has incorporated image data alongside textual features, helping in identifying misinformation more effectively. Qian et al. (2021) employed a co-attention approach to enhance text and visual characters mutually and fused the output information of every four layers of BERT with image information. Yang et al. (2019) proposed a dual-stream attention mechanism for target location perception, which can better acquire contextual information. Huang et al. (2023) modeled spatial and temporal structures to capture information dissemination and proposed a rumor detection method named STS-NN. (spatial–temporal structure neural network). The STS-NN model consists of three components: spatial capturer, temporal capturer, and integrator. All three components share parameters, allowing them to work together efficiently to identify rumors based on information dissemination. Lv et al. (2023) introduced a transformer-based model that employs an end-to-end approach to fuse multimodal feature representations into the same data domain. The model effectively captures dependencies across multiple levels of multimodal content while mitigating the impact of differences in multimodal heterogeneity.

Wan et al. (2023) developed a method involving sliding intervals to efficiently intercept necessary data instead of processing the entire sequence. To address hyper-parameter selection issues arising from integrating multiple optimization objectives, convex optimization techniques were employed to avoid excessive computational costs associated with enumeration. Throughout the training process, detection time, ACC, and stability were continuously adjusted and optimized as training objectives, enhancing the model's adaptability and generalizability. H. Li, Huang, et al. (2023) adopted

Figure 1. Overall framework of the GMRD model



bidirectional LSTM (Bi-LSTM) to extract user and text features and employed GCN to extract high-order propagation features. The complementary and alignment relationships between different features were also considered to achieve better fusion. S. Li, Wang, et al. (2023) utilized a dynamic graph attention network to encode temporal knowledge structures and an adaptive spatio-temporal and knowledge fusion network. Adaptive aggregation of knowledge information enables better integration of propagation structure information and knowledge structure information.

In recent years, several studies have integrated external knowledge into the realm of rumor detection research. Hu et al. (2021) compared social media articles with knowledge bases through entities to detect rumors. Unlike previous rumor detection models, Zheng et al. (2023) considered the role of social circles. They combined information about social circles and user behavioral preferences to create new features. These features along with the analysis of social interactions reveal clear differences between rumor sources and non-rumor sources. Gao et al. (2023) enriched post representations by introducing an auxiliary self-supervised task. They designed cluster-based and instance-based methods to analyze the relationships between various information sources. This allows the model to capture the nuances of these connections and ultimately improves its ability to identify rumors. Pi et al. (2023) proposed the PN-KG2REC (Early Rumor Detection Method Based on Knowledge Graph Representation Learning) algorithm for obtaining representations of entities and relationships. This method allows the model to turn entities and the relationships between them into a format that the rumor detection model can easily understand.

It can be observed that using both textual and image information for rumor detection can improve effectiveness. The recently popular KG can complement textual information. Therefore, this paper introduces KGs while simultaneously using textual and image features. Moreover, effective rumor detection can be achieved by effectively extracting and integrating this feature information.

Framework Overview

Figure 1 shows the overall framework of the proposed GMRD model consisting of two key components: textual semantic modeling and image modeling. Textual semantic modeling consists of two parts: extracting features from the KG using GCN and extracting features from text using multi-head attention networks and the GCN. These features are then fed into an interaction attention network. The image feature extraction layer utilizes the ResNet101 method (Q. Zhang, 2022) to extract visual features. The multimodal fusion layer combines features from text and images to obtain a more comprehensive and enriched representation.

Textual Semantic Modeling

Embedding

This layer is utilized to convert textual data into compact vector representations. Consider a dataset containing N text samples, where each text sample is represented by a sequence of words denoted as $[w_1, w_2, ..., w_T]$, with T being the sequence length. Each word w_i can be mapped to a d-dimensional word embedding vector $e_i \in \mathbb{R}^d$, thus representing the entire text sequence as a matrix $X = [e_1, e_2, \ldots, e_T]$, where $X \in \mathbb{R}^{T \times d}$.

KG

By utilizing the external KG ConceptNet (Speer et al., 2017), the model can obtain information about the meaning of each word (i.e., the semantic associations between words). This helps enrich the semantic representation of words and provides more contextual information. Based on the acquired concept information, a small-scale KG for each word can be constructed. This map includes relationships between the word and its related concepts. Such a KG can comprehensively express the semantic information of words, providing more contextual information for subsequent feature aggregation.

Entity linking method links ambiguous entities w_i in the text to the correct entities in ConceptNet. For each entity, its concept information is obtained from ConceptNet through conceptualization and represented as $C(w_i)$. Subsequently, a small-scale KG representation is constructed for each word, including concept information and word embeddings, represented in Eq. (1):

$$KG(w_i) = \left\{ C(w_i), e_i \right\} \tag{1}$$

Next, the KG representations of all words are integrated to obtain the KG representation of the entire text, as shown in Eq. (2):

$$KG(T) = KG(w_1), KG(w_2), ..., KG(w_n) \tag{2}$$

The GCNs are then utilized to aggregate the KG representation of the entire text, resulting in the text representation T_e. This is shown in Eq. (3):

$$T_e = \mathrm{GCN}(KG(T)) = \sigma\left(\widehat{D}^{-\frac{1}{2}}\widehat{A}\,\widehat{D}^{-\frac{1}{2}}X\mathbf{W}\right) \tag{3}$$

where $\widehat{A} = A + I$, A, $\widehat{D}$, and $\mathbf{W}$ are the adjacency, degree, and weight matrices, respectively, and σ is the activation function.

The Fusion Layer Combines the Multi-Head Attention Network With the GCN Network

To fully leverage the complex correlations and semantic information in text, the multi-head attention mechanism prioritizes the most crucial parts of the text sequence. Additionally, the GCN network propagates information and aggregates features within the graph structure. This allows the GCN to effectively leverage the important information extracted by the multi-head attention network. This combination enables the model to achieve a deeper understanding of the semantic meaning and contextual information of text data. This comprehensive analysis empowers the model to more accurately distinguish between rumors and non-rumors.

For the text sequence $[w_1, w_2, ..., w_T]$, each word w_i is first mapped to a word embedding vector $e_i \in \mathbb{R}^d$. Then, a multi-head attention network is constructed that can learn different attention weights

to focus on important information at different positions in the sequence. Specifically, the output of the multi-head attention network is used as the adjacency matrix for the GCN network.

For H attention heads, the attention weight matrix produced by each head is represented as $A^{(h)} \in \mathbb{R}^{T\times T}$, shown in Eq. (4):

$$A^{(h)} = \text{softmax}\left(e_i^T W^{(h)} e_j\right) \tag{4}$$

where $W^{(h)} \in \mathbb{R}^{d\times d}$ is the parameter matrix for each attention head, $h \in [1, H]$.

Then, these attention weight matrices are concatenated column-wise to form a matrix of size $T \times (HT)$, denoted as $A^{(att)}$. This is shown in Eq. (5):

$$A^{(att)} = \left[A^{(1)}, A^{(2)}, \ldots, A^{(H)}\right] \tag{5}$$

Next, $A^{(att)}$ is normalized to obtain the final attention adjacency matrix $\widehat{A}^{(att)}$, as shown in Eq. (6):

$$\widehat{A}^{(att)} = \text{normalize}(A^{(att)}) \tag{6}$$

Finally, the attention adjacency matrix is processed using the GCN network to obtain the text feature representation, as shown in Eq. (7):

$$Y = GCN\left(\widehat{A}^{(att)} X W_{gcn}\right) \tag{7}$$

Interaction Attention Network Fusion (INT-Att)

The results from the KG layer are used as the query vectors for attention, while the multi-head attention and GCN network layer results are used as key-value vectors. This approach lets the model focus on crucial rumor-related content within the text by selectively attending to relevant information from the KG. It enables the model to adaptively attend to relevant content in the text and combine it with semantic information from the KG for weighted fusion, thereby enhancing the detection capability of rumors.

First, the attention weights $\alpha^{T_e,Y}$ are calculated using Y and text features T_e, as shown in Eq. (8):

$$\alpha^{T_e,Y} = \text{softmax}\left(T_e W_q \left(Y W_k\right)^T\right) \tag{8}$$

where W_q and W_k are the parameter matrices of T_e, and Y, respectively.

Next, attention weights are used to compute a weighted sum of text features T_e, obtaining the interaction feature Z, as shown in Eq. (9):

$$Z = \alpha^{T_e,Y} Y \tag{9}$$

Image Information Extraction Layer

The image feature processing section utilizes ResNet101 to process image features, obtaining the image feature vector f. ResNet101 is a deep CNN commonly used for image recognition and feature extraction tasks. Given an input image I, ResNet101 maps it to a feature vector f, which can be represented as a function. This is shown in Eq. (10):

$$f = \text{ResNet101}(I) \tag{10}$$

where f is the image feature vector extracted by ResNet101. This process typically involves multiple convolutional and pooling layers, ultimately mapping the image to a fixed-dimensional feature vector. While the exact details of the network architecture and parameters are beyond the scope of this paper, the use of a powerful extractor like ResNet101 ensures the extraction of image features with rich semantic information.

Prediction

The text interaction feature Z and the image feature vector f are concatenated and then fused through a fully connected layer to predict rumors, as shown in Eq. (11) and Eq. (12):

$$C = [Z,f] \tag{11}$$

$$W_{out} = \sigma\left(C \cdot W_{fc} + b_{fc}\right) \tag{12}$$

where W_{fc} is the parameter matrix, b_{fc} is the bias, and σ is the activation function. Therefore, the output of the fully connected layer can be represented as W_{out}.

Optimization

The model was trained using cross-entropy loss function, as shown in Eq. (13):

$$\mathscr{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\left(\hat{y}_{i,c}\right) \tag{13}$$

where N is the number of samples, C is the number of rumor classes, $y_{i,c}$ is the true label for the i-th sample corresponding to the c-th rumor class, and $\hat{y}_{i,c}$ is the predicted label.

```
Algorithm 1. GMRD
1: Initialize:
Weights, bias parameters
Begin:
2: For epoch in n do
3: Get the embedding vectors X
4: Get the KG representation KG(T)
5: Get the GCN representation Te
6: Get the text representation Y
7: Get the final text representation Z
8: Get the image representation f
9: Compute the rumor prediction label  Wout
10: Update parameters
11: End For
```

Experimental Results

To assess the effectiveness of the proposed GMRD model, a series of experiments were conducted. The performance of the proposed model was compared against established baseline methods on publicly available datasets. This evaluation aimed to answer the following three questions:

Table 1. The statistical information of the twitter and Weibo datasets

	Rumor	Non-Rumor	Total
Weibo	4,635	4,723	9,358
Twitter	7,256	5,837	13,093

1. How does the performance of the GMRD method compare to state-of-the-art methods?
2. What is the impact of key model designs of GMRD on the experimental results?
3. How do hyper-parameter settings affect the experimental results?

Datasets

Table 1 shows the statistical information of the Twitter and Weibo datasets.

To evaluate the effectiveness of GMRD and avoid experimental contingencies, experiments were conducted on two publicly available datasets as shown in Table 1: Twitter (Maigrot et al., 2016) and Weibo. The Twitter dataset consists of tweets, each containing textual content, images/videos, and relevant social context information. Since this study focuses only on textual and image information, tweets with videos were excluded from both datasets. Additionally, tweets lacking text or images were also not included in the analysis. Within the Weibo dataset, authentic data undergoes verification by the credible Chinese news agency Xinhua News Agency, whereas fabricated news is authenticated through Weibo's official debunking system. Each tweet in the dataset includes text, images, videos, and social attributes. The dataset was divided into training, validation, and testing sets in a 7:1:2 ratio.

Evaluation Metrics

Evaluation of the overall performance of the model was conducted using the evaluation metrics ACC and F1 score. The formulas for calculation are shown in Eq. (14), Eq. (15), Eq. (16), and Eq. (17):

$$Acc = \frac{(TP + TN)}{N} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1 = \frac{2 \times Precision \times Recall}{recision + Recall} \tag{17}$$

TP (True Positive) denotes the count of samples accurately classified as a specific rumor category by the rumor classifier. TN (True Negative) denotes the count of samples accurately classified as non-members of a particular rumor category. FP (False Positive) denotes the count of samples incorrectly classified as belonging to a specific rumor category. FN (False Negative) denotes the count of samples incorrectly classified as not belonging to a specific rumor category.

Baselines

1. Only-Text: For the only-text setting, the GMRD model only uses text features for rumor detection without utilizing image features.
2. Only-Image: In the only-image configuration, the GMRD model utilizes only image features. Text features are entirely disregarded. Visual features are obtained by feeding into the ResNet101 model for rumor detection.
3. ATT-RNN (Recurrent Neural Network with an Attention Mechanism) (Jin et al., 2017): It is a model based on attention mechanism and RNN for modeling and predicting sequential data tasks. ATT-RNN performs rumor detection after integrating features from text, images, and social context. For comparison with the proposed model, social attributes in tweets were removed.
4. EANN (Event Adversarial Neural Networks) (Wang et al., 2018): This model utilizes an event discriminator to eliminate event-specific features. It learns transferable features for rumor detection concerning newly emerging occurrences and time-critical events.
5. MVAE (Multimodal Variational Autoencoder For Fake News Detection) (Khattar et al., 2019): It is a novel end-to-end multimodal variational autoencoder, which utilizes textual and visual information for rumor detection. The processing presented in the original text was followed without alterations in the experiments of this study.
6. CARMN (Crossmodal Attention Residual and Multichannel convolutional neural Networks) (Song et al., 2021): In this model, textual feature representations are extracted from raw text and fused text using an (Multichannel Convolutional Neural Network) MCN, while image feature representations are extracted using VGG19 for multimodal rumor detection.
7. CAFE (Cross-modal Ambiguity Learning) (Y. Chen et al., 2022): This model analyzes cross-modal fuzzy learning from an information-theoretic perspective and performs adaptive aggregation of uni-modal and cross-modal correlated features.
8. TDEDA (Dual-Attention Based on Textual Double Embedding) (Han et al., 2023): It is a neural network for multimodal fusion in rumor detection that facilitates enhanced information exchange at the level of text-image objects. It leverages an attention mechanism to capture visual features associated with keywords.

Hyper-Parameter

The parameter settings of this experiment refer to previous work (Jin et al., 2017). The learning rate was set to 2e-5, the number of attention heads was eight, the maximum text length was set to 40, and the word embedding dimension was 100. The image input format was adjusted to 224*224*3. An Adam optimizer was used with a batch size of 128. Cross-entropy was employed as the loss function with a dropout rate of 0.5. The GCN consisted of two layers.

Comparative Analysis (RQ1)

Analysis of the two datasets reveals that the text-based GMRD model consistently outperformed the image-only model. This suggests that text information played a more prominent role than visual information in the conducted experiments. Furthermore, performance metrics on the Weibo dataset exhibit superior results compared to those on the Twitter dataset for both text and image models. This difference can be attributed to the varying habits of Chinese and international netizens, with Weibo posts generally being longer than Twitter tweets, resulting in better performance of GCN on the Weibo dataset. However, the performance is still not as good as that of multimodal information. Furthermore, the CARMN method, which leverages a multi-channel CNN to detect rumors using both image and text features, achieves superior results. This finding supports the notion that using multimodal features is a promising approach for rumor detection. Additionally, multimodal aggregation

Table 2. Comparative analysis

Dataset	Method	ACC	Rumor			Non-Rumor		
			Precision	Recall	F1	Precision	Recall	F1
Twitter	Only-Text	0.704	0.667	0.542	0.598	0.712	0.633	0.670
	Only-Image	0.601	0.693	0.532	0.602	0.529	0.693	0.600
	ATT-RNN	0.686	0.786	0.631	0.700	0.613	0.778	0.686
	EANN	0.723	0.65	0.481	0.553	0.553	0.769	0.643
	CARMN	0.743	0.853	0.625	0.721	0.681	0.883	0.769
	CAFE	0.809	0.813	0.792	0.802	0.806	0.82	0.813
	TDEDA	0.827	0.781	0.853	0.815	0.867	0.812	0.839
	GMRD	**0.843**	**0.787**	**0.871**	**0.827**	**0.854**	**0.834**	**0.844**
Weibo	Only-Text	0.806	0.811	0.853	0.831	0.843	0.765	0.802
	Only-Image	0.632	0.64	0.543	0.588	0.635	0.763	0.693
	ATT-RNN	0.789	0.858	0.682	0.760	0.767	0.887	0.823
	EANN	0.813	0.825	0.821	0.823	0.814	0.809	0.811
	CARMN	0.856	0.897	0.816	0.855	0.821	0.896	0.857
	CAFE	0.847	0.853	0.827	0.840	0.821	0.856	0.838
	TDEDA	0.877	0.849	0.841	0.845	0.895	0.893	0.894
	GMRD	**0.881**	**0.856**	**0.848**	**0.852**	**0.899**	**0.901**	**0.900**

is also crucial. The CAFÉ method proposes aggregating uni-modal and cross-modal features from an information theory perspective, leading to significantly improved detection performance. By considering both image and text features, and employing dual self-attention mechanisms to capture richer internal feature information, the model fully explores the potential connections between text and image objects.

To enrich the understanding of textual features, the GMRD leverages KGs. In the text modeling stage, a multi-head attention network and the GCN are used. The interaction attention network effectively connects the external knowledge from the KG with the information extracted from the text features. Furthermore, the ResNet101 network is utilized to extract image features, resulting in significant improvements in performance compared to other methods. Table 2 shows the comparative analysis.

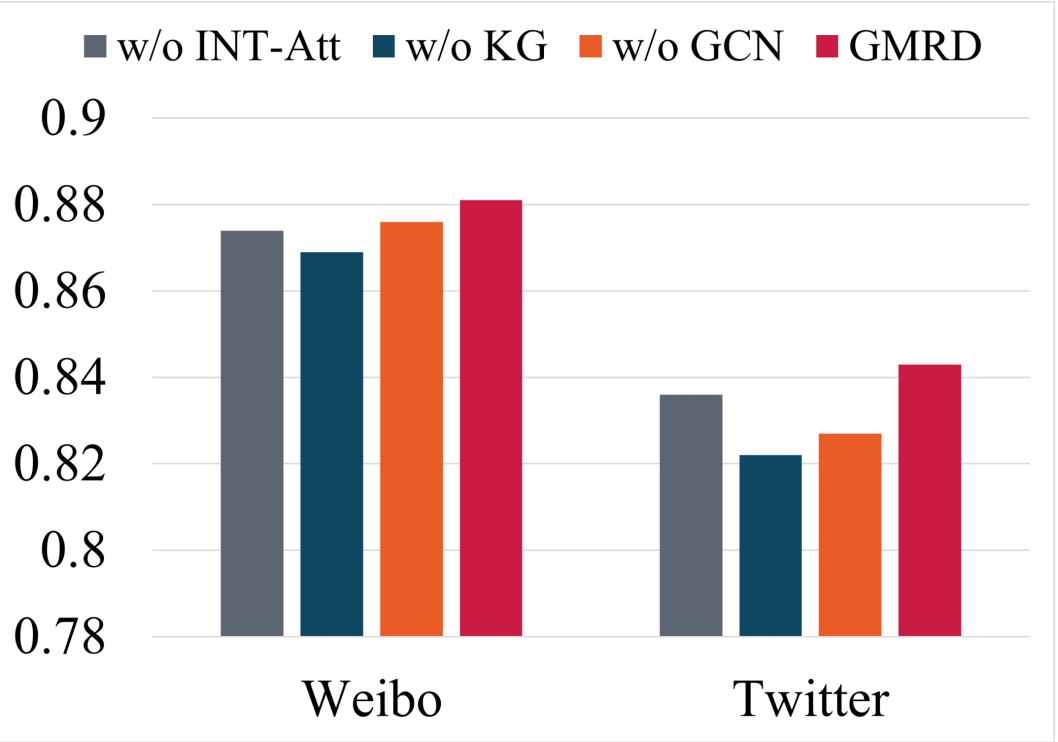
Ablation (RQ2)

To investigate the influence of different components of GMRD on the overall performance, separate tests were conducted for the INT-Att, the KG, and the GCN network.

Effectiveness of the GCN

In Figure 2, "w/o GCN" represents the scenario where GCN is not utilized. It can be observed that the performance is inferior without GCN. This is because GCN possesses a multi-layer network structure, allowing for the gradual extraction and fusion of multi-level feature representations in textual data through multi-layer information propagation and aggregation. Consequently, this enhances the model's capability to extract rumor-related features. By processing information locally within the graph structure, the GCN allows the model to be more flexible and adaptable. This means the model can handle different types of textual data effectively. Therefore, by incorporating GCN, a

Figure 2. Effectiveness of different components



more comprehensive exploration of information in textual data is achieved, leading to a significant improvement in the model's ability to detect rumors.

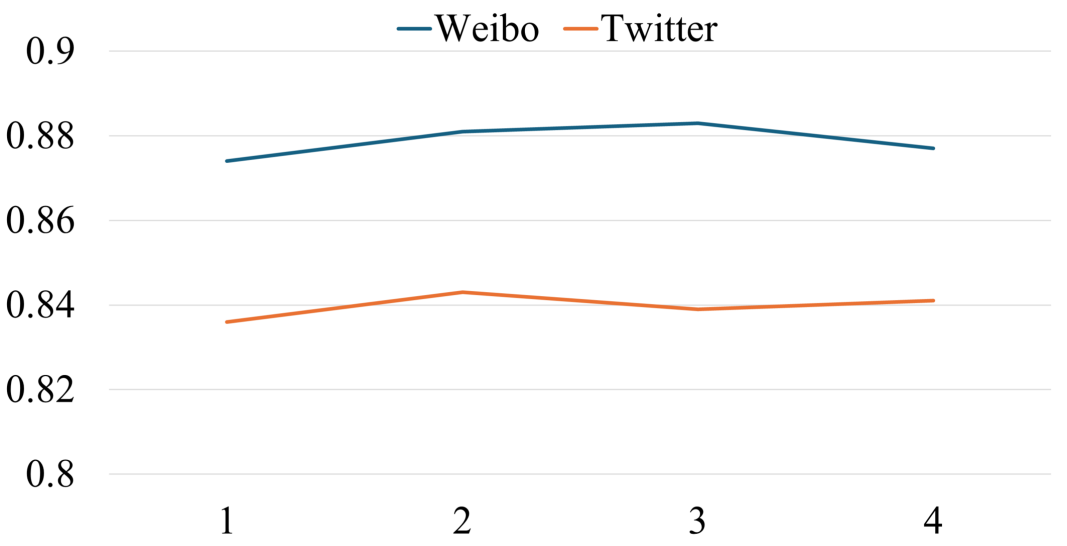
Effectiveness of the KG

In Figure 2, "w/o KG" denotes the scenario where the KG was not utilized as an external feature. It can be seen from Figure 2 that the performance is inferior compared to GMRD when the KG is not incorporated. The reason lies in the fact that rumor detection tasks require a thorough understanding of textual data, including semantic information and entity relationships. ConceptNet, as an open-domain KG, contains a vast amount of common knowledge and entity relationships, providing the model with rich semantic information and entity associations to better comprehend textual content. Furthermore, ConceptNet aids in constructing semantic associations, connecting entities and concepts in the text to form a more comprehensive and accurate semantic network. This network allows the model to capture implicit information and semantic relationships in textual data. By understanding these deeper connections, the model becomes more effective at identifying rumors in the text data. Additionally, ConceptNet facilitates cross-domain knowledge transfer, enabling the model to enrich textual feature representations with knowledge from different domains, improving the model's generalization and adaptability. Thus, the incorporation of the ConceptNet KG provides the model with richer semantic information and entity associations, resulting in improved rumor detection and classification performance.

Effectiveness of the Interaction Attention Fusion Network

In Figure 2, "w/o INT-Att" denotes the scenario where the interaction attention fusion network was not present. It can be seen from Figure 2 that the GMRD model performs better, attributed to the

Figure 3. Impact of the number of GCN layers



fact that the task of rumor detection necessitates the integration of multimodal information such as text and images. The interaction attention fusion network effectively merges these two modalities and leverages their correlation. By introducing attention mechanisms, the network autonomously learns the relational dynamics between text and image. This enhances the model's ability to comprehend semantic relationships between textual content and image features, ultimately leading to improved ACC in rumor identification.

Moreover, the interaction attention fusion network facilitates information interaction and propagation across different feature spaces. This allows the model to take advantage of the unique strengths of both text and image features, resulting in a richer and more comprehensive understanding of the data. Additionally, the network effectively handles heterogeneity and varying scales of feature representations across different modalities, making the model more robust and generalizable.

Thus, incorporating the interaction attention fusion network enables better integration of multimodal information such as text and images, enhancing the effectiveness of rumor detection tasks.

Hyperparameter Analysis (RQ3)

In this experiment, the number of GCN layers was set to [1, 2, 3, 4]. Figure 3 shows that the performance first increases and then decreases. This is because shallower networks (with one or two layers) fail to capture intricate connections within the textual features, which limits their ACC. On the other hand, deeper networks (with three or four layers) may suffer from issues such as vanishing gradients and over-fitting, resulting in reduced generalization performance. When the number of layers is two, the model can adequately capture the relationships among textual features at a moderate complexity level, improving the representational and generalization capabilities of the GMRD. However, adding even more layers can become counterproductive. The model becomes increasingly complex, making it harder to train effectively and increasing the risk of getting stuck in local optima.

DISCUSSION

The strengths of the proposed model lie in several aspects. Firstly, the model leverages the ConceptNet to construct small-scale KGs. It then utilizes GCN to effectively capture semantic

Table 3. Abbreviations

Abbreviation	Full Form
DNN	Deep-Learning Neural Network
CNN	Convolutional Neural Networks
NLP	Natural Language Processing
GCN	Graph Convolutional Network
KG	Knowledge Graph
INT-Att	Interaction Attention Network Fusion
Bi-LSTM	Bidirectional Long Short-Term Memory Networks
ACC	Accuracy

information and entity relationships in text data. This combined approach significantly improves the model's ability to detect rumors. Secondly, by introducing the interactive attention fusion network, the model interacts with and integrates text and image features. This allows the model to fully exploit the correlation between multimodal information and increase the model's ACC in discriminating rumors.

However, the model also has some limitations. Firstly, it involves multiple complex computational steps, including multi-head attention networks, KG construction, and the ResNet101 method. These steps lead to high computational complexity and increase the training and inference times. Secondly, the model relies on external data sources such as the ConceptNet KG, which has a high dependence on the quality and update frequency of the data and may be affected by changes in external data. Although the process fusion method effectively integrates text and image information, optimizing its fusion strategy and parameter selection holds promise for further enhancing performance and stability.

CONCLUSION

This paper proposes a multimodal rumor detection model that integrates text and image information. The model leverages multi-head attention networks, KGs, the ResNet101 method, and process fusion strategies to effectively integrate textual and visual information and achieve high ACC. Future work will include further optimizing the proposed multimodal rumor detection model, including adjusting parameters, improving feature extraction methods, and optimizing fusion strategies to enhance the model's performance and effectiveness. The model's applications can be broadened by investigating its effectiveness in new domains such as rumor identification and rumor generation. Additionally, research is needed to validate its performance across diverse languages and cultural backgrounds.

AUTHOR NOTE

The authors declare no conflicts of interest.

The data used to support the findings of this study are included within the article.

This research was supported by The University Key Scientific Research Project of Henan Province (no. 22A520052). The University Key Scientific Research Project of Henan Province (no. 24B880075).

Correspondence concerning this article should be addressed to Qian Li, School of Computer Science and Technology, Zhoukou Normal University, ZhouKou Henan 466001, China. Email: liqian@zknu.edu.cn

PROCESSING DATES

This manuscript was initially received for consideration for the journal on 03/14/2024, revisions were received for the manuscript following the double-anonymized peer review on 05/30/2024, the manuscript was formally accepted on 05/21/2024, and the manuscript was finalized for publication on 05/31/2024

REFERENCES

Bai, N., Meng, F., Rui, X., & Wang, Z. (2021). Rumour detection based on graph convolutional neural net. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 21686–21693. 10.1109/ACCESS.2021.3050563

Bin, S., & Sun, G. (2022). Research on the influence maximization problem in social networks based on the multi-functional complex networks model. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(3), 1–17. 10.4018/JOEUC.302662

Chen, T., Wu, L., Li, X., Zhang, J., Yin, H., & Wang, Y. (2018, June 3). *Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection* [Conference presentation]. Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2018 Workshops, BDASC, BDM, ML4Cyber, PAISI, DaMEMO, Melbourne, VIC, Australia.

Chen, Y., Li, D., Zhang, P., Sui, J., Lv, Q., Tun, L., & Shang, L. (2022). Cross-modal ambiguity learning for multimodal fake news detection. *InProceedings of the ACM web conference 2022*, 2897-2905. 10.1145/3485447.3511968

Dun, Y., Tu, K., Chen, C., Hou, C., & Yuan, X. (2021). *Kan: Knowledge-aware attention network for fake news detection* [Conference presentation]. Proceedings of the AAAI Conference on Artificial Intelligence.

Gao, Y., Wang, X., He, X., Feng, H., & Zhang, Y. (2023). Rumor detection with self-supervised learning on texts and social graph. *Frontiers of Computer Science*, *17*(4), 174611. 10.1007/s11704-022-1531-9

Guo, Z., Yu, K., Jolfaei, A., Bashir, A. K., Almagrabi, A. O., & Kumar, N. (2021). Fuzzy detection system for rumors through explainable adaptive learning. *IEEE Transactions on Fuzzy Systems*, *29*(12), 3650–3664. 10.1109/TFUZZ.2021.3052109

Han, H., Ke, Z., Nie, X., Dai, L., & Slamu, W. (2023). Multimodal fusion with dual-attention based on textual double-embedding networks for rumor detection. *Applied Sciences (Basel, Switzerland)*, *13*(8), 4886. 10.3390/app13084886

Hu, L., Yang, T., Zhang, L., Zhong, W., Tang, D., Shi, C., Duan, N., & Zhou, M. (2021). *Compare to the knowledge: Graph neural fake news detection with external knowledge* [Conference presentation]. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 1.

Huang, Q., Zhou, C., Wu, J., Liu, L., & Wang, B. (2023). Deep spatial–temporal structure learning for rumor detection on Twitter. *Neural Computing & Applications*, *35*(18), 12995–13005. 10.1007/s00521-020-05236-4

Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017). *Multimodal fusion with recurrent neural networks for rumor Ddtection on microblogs* [Conference presentation]. Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, United States.

Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). *MVAE: Multimodal variational autoencoder for fake news detection* [Conference presentation]. The World Wide Web Conference, San Francisco, CA, United States.

Lawson-Body, A., Jackson, J., Hinsz, V., Illia, A., & Lawson-Body, L. (2023). Cybersecurity and social media networks for donations: An empirical investigation of triad of trust, commitment, and loyalty. [JOEUC]. *Journal of Organizational and End User Computing*, *35*(1), 1–26. 10.4018/JOEUC.332062

Li, H., Huang, G., Li, C., Li, J., & Wang, Y. (2023). Adaptive spatial–temporal and knowledge fusing for social media rumor detection. *Electronics (Basel)*, *12*(16), 3457. 10.3390/electronics12163457

Li, S., Wang, Y., Huang, H., & Zhou, Y. (2023). Study on the rumor detection of social media in disaster based on multi-feature fusion method. *Natural Hazards*, ●●●, 1–20. 10.1007/s11069-023-06005-x

Liao, X., Zhi, H., Yang, D., Cheng, X., & Chen, G. (2018). Rumor detection in social media based on hierarchical attention network. *Sci Sin Inform*, *48*(11), 1558–1574. 10.1360/N112018-00134

Lin, D., Ma, B., Cao, D., & Li, S. (2019). Chinese microblog rumor detection based on deep sequence context. *Concurrency and Computation*, *31*(23), e4508. 10.1002/cpe.4508

Liu, Z., Wei, Z. H., & Zhang, R. X. (2017). Rumor detection based on convolutional neural network. *Jisuanji Yingyong*, *37*(11), 3053.

Ly,Wang, X., & Shao, C. (2023). TMIF: Transformer-based multi-modal interactive fusion for automatic rumor detection. *Multimedia Systems*, *29*(5), 2979–2989. 10.1007/s00530-022-00916-8

Ma, J., Gao, W., & Wong, K. (2018). *Rumor detection on twitter with tree-structured recursive neural networks*. Association for Computational Linguistics. 10.18653/v1/P18-1184

Maigrot, C., Claveau, V., Kijak, E., & Sicre, R. (2016). MediaEval 2016: A multimodal system for the verifying multimedia use task. In *MediaEval 2016: "Verfiying multimedia use" task*.

Peng, Y., & Wang, J. (2021). Rumor detection based on attention CNN and time series of context information. *Future Internet*, *13*(11), 267. 10.3390/fi13110267

Pi, D. C., Wu, Z. Y., & Cao, J. J. (2023). Early rumor detection method based on knowledge graph representation learning. *Acta ElectonIca Sinica*, *51*(2), 385.

Qian, S., Wang, J., Hu, J., Fang, Q., & Xu, C. (2021, July). Hierarchical multi-modal contextual attention network for fake news detection. *InProceedings of the 44th international ACM SIGIR conference on research and development in information retrieval,* 153-162. https://doi.org/10.1145/3404835.3462871

Qin, Q., Yang, X., Zhang, R., Liu, M., & Ma, Y. (2022). An application of deep belief networks in early warning for cerebrovascular disease risk. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(4), 1–14. 10.4018/JOEUC.287574

Singhal, S., Shah, R. R., Chakraborty, T., Kumaraguru, P., & Satoh, S. I. (2019). Spotfake: A multi-modal framework for fake news detection. *In 2019 IEEE fifth international conference on multimedia big data (BigMM)*, 39-47. 10.1109/BigMM.2019.00-44

Song, C., Ning, N., Zhang, Y., & Wu, B. (2021). A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. *Information Processing & Management*, *58*(1), 102437. 10.1016/j.ipm.2020.10243733041437

Speer, R., Chin, J., & Havasi, C. (2017). Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the AAAI conference on artificial intelligence* (*Vol. 31*, No. 1). https://doi.org/10.1609/aaai.v31i1.11164

Wan, P., Wang, X., Pang, G., Wang, L., & Min, G. (2023). A novel rumor detection with multi-objective loss functions in online social networks. *Expert Systems with Applications*, *213*, 119239. 10.1016/j.eswa.2022.11923936407849

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., & Gao, J. (2018). *EANN: Event adversarial neural networks for multi-modal fake news detection* [Conference presentation]. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, United Kingdom.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J. G., Salakhutdinov, R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, ●●●, 32.

Zhang, K., Yu, J., Shi, H., Liang, J., & Zhang, X. (2023). *Rumor detection with diverse counterfactual evidence* [Conference presentation]. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.

Zhang, Q. (2022). A novel ResNet101 model based on dense dilated convolution for image classification. *SN Applied Sciences*, *4*(1), 1–13. 10.1007/s42452-021-04897-7

Zheng, P., Huang, Z., Dou, Y., & Yan, Y. (2023). Rumor detection on social media through mining the social circles with high homogeneity. *Information Sciences*, *642*, 119083. 10.1016/j.ins.2023.119083

Zhou, Z., Qi, Y., Liu, Z., Yu, C., & Wei, Z. (2018). *A C-GRU neural network for rumors detection* [Conference presentation]. 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS).

Qian Li, Associate Professor, Master's degree, Graduated from Huazhong University of Science and Technology in 2007. Worked in Zhoukou Normal University. His research interests include Network security, Internet of things security.

Laihang Yu, Associate Professor, Doctor's degree, Graduated from Dalian University of Technology in 2018. Worked in Zhoukou Normal University. His research interests include Network security, Image Processing.

Pan Li, graduated from Huazhong University of Science and Technology in December 2008, bachelor's degree (July 2005), master's degree, associate professor title, teacher of Zhengzhou Institute of Engineering and Technology, big data, cloud computing 1https://twitter.com.