IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 20 May 2023, accepted 6 June 2023, date of publication 8 June 2023, date of current version 14 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3284308

RESEARCH ARTICLE

A Novel Rumor Detection Method Based on Non-Consecutive Semantic Features and Comment Stance

YI ZHU[1], GENSHENG WANG[2,3], SHENG LI[1], AND XUEJIAN HUANG[3]

[1]School of Finance, Taxation and Public Administration, Jiangxi University of Finance and Economics, Nanchang 330013, China
[2]School of International Economics and Trade, Jiangxi University of Finance and Economics, Nanjing 330013, China
[3]School of Humanities, Jiangxi University of Finance and Economics, Nanchang 330013, China

Corresponding author: Gensheng Wang (wgs74@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 72061015, and in part by the Science and Technology Project of Jiangxi Provincial Department of Education under Grant GJJ200539.

ABSTRACT Detecting rumors on social media has become increasingly necessary due to their rapid spread and adverse impact on society. Currently, most rumor detection methods fail to consider the non-consecutive semantic features of the post's text or the authority of commenting users. Therefore, we propose a novel rumor detection method that integrates non-consecutive semantic features and stances considering user weights. Firstly, we employ a pre-trained stance detection model to extract stance information for each comment for the post and then determine the weight of the stance information based on commenting user characteristics. Secondly, we input the time-series data of stance information and the corresponding comment user sequence data into the Cross-modal Transformer to learn the temporal features of comment stances. We then use pointwise mutual information to transform the discretized and fragmented post's text into a weighted graph and utilize a graph attention network that considers edge weights to process the graph and learn the non-consecutive semantic features of the text. Finally, we concatenate the temporal features of comment stances with the non-consecutive semantic features of the post's text and input them into a multi-layer perceptron for rumor classification. Experimental results on two public social media rumor datasets, Weibo and PHEME, demonstrate that our method outperforms the baselines. Our method is at least 12 hours ahead of the baseline methods for early rumor detection.

INDEX TERMS Rumor detection, non-consecutive semantic features, weighted graph attention network, cross-modal transformer, attention mechanism, deep learning.

I. INTRODUCTION

With the rapid advancement of the Internet and mobile technology, social media platforms such as Twitter, Facebook, and Weibo have become crucial channels for people to access and share information. However, while social media promotes information exchange, it also fosters a new breeding ground for rumor spreading. The dissemination of rumors hinders the effective use of social media and may cause misunderstandings among the public, trigger negative emotions, and disrupt social order [1]. For example, during the early stages of the COVID-19 pandemic, many rumors circulated on social media, such as "COVID-19 vaccines will alter human DNA," "5G networks caused the spread of the virus," and "COVID-19 virus is man-made," which led to public misunderstandings and panic. In order to curb the spread of rumors, some organizations have established information verification outlets, such as FactCheck[1], Politifact[2], and Snopes[3]. Nevertheless, these outlets mainly depend on manual validation, which is labor-intensive and has a long delay in debunking rumors. Therefore, studying the

The associate editor coordinating the review of this manuscript and approving it for publication was Ines Domingues.

[1]https://www.factcheck.org
[2]https://www.politifact.com
[3]https://www.snopes.com

automatic detection method of social media rumors and effectively debunking rumors during their incubation period is significant.

Previous rumor detection methods mainly rely on statistical machine learning models based on manual feature engineering. The research focuses on designing rumor features guided by prior knowledge, but handcrafted features lack comprehensiveness and flexibility to represent the deep semantic features of rumors. With the development of deep learning technology, researchers have started using deep learning models to learn the deep semantic features of rumors automatically. Rumor detection has gradually entered the data-driven era. Due to length restrictions on social media platforms, posts are typically concise and often structured as "#hashtags# text @user #text [emoji] [symbol]...". The brevity of posts and the prevalent use of abbreviations, symbols, and emojis contribute to the fragmented and discrete nature of textual content. Currently, deep learning approaches for rumor detection based on content features mainly focus on the consecutive semantic features of rumors and do not consider social media posts' discrete and fragmented nature. Furthermore, sometimes rumor spreaders deliberately mimic the writing style of factual information, leading to situations where methods relying solely on content features may fail to perform well. Deep learning methods based on propagation features for rumor detection focus on rumors' propagation patterns or the retweets' comments. However, most methods do not distinguish the weights of different comment users. Intuitively, we should pay more attention to the comments of authoritative users.

Therefore, we propose a novel rumor detection method that integrates non-consecutive semantic features and users' stances and features, called SFCR, which implies that rumor detection primarily relies on semantic features and comments. First, we use a pre-trained stance detection model to extract stance information for each comment and then determine the stance information's weight according to the commenting user's features. Second, we feed the time-series data of stance information and the corresponding sequence data of comment users into the Cross-modal Transformer to learn the temporal features of comment stances. Then, we convert the discretized and fragmented post's text into weighted graphs using pointwise mutual information and employ a graph attention network (GAT) that considers edge weights to learn non-consecutive semantic features of the texts. Finally, we integrate the temporal features of stances and the non-consecutive semantic features for rumor classification. Experimental results on two public social media rumor datasets, Weibo and PHEME, show that our method outperforms the baseline methods. The main contributions of this paper are as follows:

1. We propose a novel rumor detection method that integrates non-consecutive semantic features of text and users' stances in comments.

2. We propose a method for learning non-consecutive semantic features of the post's text using a weighted graph attention network.

3. We propose a comment stance learning method that considers user features, which pay more attention to the comments of authoritative users and thus avoids the deliberate hype of internet trolls.

4. We conduct comprehensive experiments on two public datasets, demonstrating that our method outperforms the baseline methods.

II. RELATED WORK

A. RUMOR DETECTION METHODS BASED ON CONTENT FEATURES

Content-based methods regard rumor detection as a text classification task. Currently, the mainstream strategy is to learn the deep semantic features of the post's text through deep learning models. For example, Kaliyar et al. [2] propose a network architecture called FNDNet based on Convolutional Neural Network (CNN), which can automatically extract helpful features for rumor classification from text content. Ajao et al. [3] combine Long Short-Term Memory Network (LSTM) and CNN to extract the semantic features of rumors on Twitter. Alkhodair et al. [4] propose a rumor detection model that combines LSTM and Word2vec, which can accurately identify emerging topic rumors based solely on the post's text. Cheng et al. [5] propose a rumor detection model based on Generative Adversarial Networks (GANs), which strengthens the learning of rumor semantic features through the mutual promotion of the discriminator and the generator. In order to improve the accuracy, some researchers have explored additional features beyond semantic ones. For instance, Xu et al. [6] introduce a topic-driven rumor detection model called TDRD, which utilizes CNN to extract the topic information from the content and combines it with Word2vec for rumor detection. Ma et al. [7] introduce a rumor detection method that obtains entity explanations through knowledge graphs, enhancing the semantic understanding of rumor texts. Singh et al. [8] integrate text semantic features extracted by Attention-based LSTM and user features for rumor detection. Kaliyar et al. [9] propose a rumor detection method that constructs a multi-dimensional tensor matrix by combining user-news relationships, news content, and user-user relationships. They then perform tensor decomposition on the matrix to acquire the fusion features of users and content and use them for rumor detection. To address the issue of single-modal fake news detection methods failing to detect fake news based on complete multi-modal information, Shao et al. [10] propose a novel method for multi-modal fake news detection, called fake news detection based on multi-modal classifier ensemble. This method takes into account the advantages of both single-modal and multi-modal methods. To address the issue of over-reliance on content features and the inability to assess individual users' impact on rumor spreading accurately, Chen et al. [11] propose a model called UMLARD. This model effectively learns the representation of diverse perspectives from users who engaged in spreading the tweet and integrates these learned features using a distinguishable fusion mechanism.

B. RUMOR DETECTION METHODS BASED ON PROPAGATION FEATURES

Propagation-based methods primarily employ features such as retweets, comments, and propagation structures during the spread of rumors for rumor detection [12], [13]. For example, Ma et al. [14] propose a propagation tree kernel (PTK) method to capture the high-order patterns of rumor propagation based on the propagation structure of rumors. Ma et al. [15] uses temporal comment data to construct a rumor detection model based on Recurrent Neural Network (RNN). Xu et al. [16] filters out the comments of users with low influence according to the number of followers, and then rumor detection is performed on the filtered data using RNN. References [17] and [18] introduce a rumor detection model based on Tree-structured Recursive Neural Networks, which simultaneously learns the structural features of rumor propagation and the semantic features of comment data. In order to make the model pay more attention to the part of the propagation data with rumor traits, Chen et al. [19] propose a rumor detection model that combines RNN with an attention mechanism. In response to the problem that current rumor detection methods based on propagation structure ignore the temporal features of propagation, Huang et al. [20] propose a spatio-temporal rumor detection model that integrates structure and temporal information. With the advancements in Graph Neural Networks (GNN), [21], [22], [23] have proposed a rumor detection model based on GNN, which learns the representation of propagation structure through GNN. References [24] and [25] have found that comments reflect the stance of other users towards the post, such as skepticism, opposition, and agreement. These comments' stances influence each other and are crucial for detecting rumors. Therefore, [26], [27], [28], [29], [30] propose a stance-based approach for rumor detection. In order to relieve the model's dependence on propagation features and improve early detection performance, Tu et al. [31] propose a rumor detection model Rumor2vec that jointly learns text and propagation structure representations. Experiments demonstrate that the method can identify rumors at least 12 hours earlier. Lotfi et al. [32] introduce a rumor detection approach that integrates propagation and user features to learn information from user interaction graphs and rumor propagation graphs through Graph Convolutional Networks (GCN). Sun et al. [33] propose a novel approach called Dual-Dynamic Graph Convolutional Networks (DDGCN), which integrates the dynamics of message propagation and the dynamics of background knowledge from Knowledge graphs into a unified framework. Xu et al. [34] present a rumor detection model, Hierarchically Aggregated Graph Neural Networks (HAGNN), which focuses on capturing different granularities of high-level representations of text content and fusing the rumor propagation structure.

Content-based rumor detection methods automatically extract the semantic features of rumors through deep learning. Some researchers [6], [7], [8], [35] have integrated topics, entity concepts,external knowledge, or user features into semantic features to improve their accuracy. However, most current content-based rumor detection methods do not consider the discrete and fragmented characteristics of the post's text and do not explore the non-consecutive semantic features of the text. Propagation-based rumor detection methods extract the temporal and structural features of rumor propagation through RNN and GNN. To improve early detection performance, some researchers [31], [32], [33], [34], [36] have integrated content features, user features, background knowledge, or social relation into propagation features. The stance of comment users towards a post plays a crucial role in rumor detection. However, current stance-based rumor detection methods do not differentiate the weight of different comment users. Therefore, we propose a novel rumor detection method integrating non-consecutive semantic features and stances considering user weights. Firstly, we employ a weighted GAT to learn non-consecutive semantic features of the post's text. Then, we utilize a Cross-modal Transformer to learn comment stance features that consider the weight of users. Finally, we integrate non-consecutive semantic features with comment stance features to detect the rumor, improving the accuracy and timeliness of the method.

III. METHODOLOGY

Given a set of posts $P = \{p_1, p_2, \ldots, p_n\}$, $p_i = (T_i, C_i)$, where $T_i = \{w_1, w_2, \ldots, w_n\}$ represents the textual information of post p_i, and $C_i = \{(u_1, t_1, c_1), (u_2, t_2, c_2), \ldots, (u_m, t_m, c_m)\}$ represents the user's comments on the post p_i. Here, u_m, t_m, and c_m represent the user, time, and content information of the m-th comment. We aim to learn a model $\int : P \rightarrow Y$ that takes the textual information T_i and the user comments C_i of a post p_i as inputs to classify p_i into pre-defined categories $Y = \{0, 1\}$, where 0 denotes non-rumor, and 1 denotes rumor. The overall framework of our proposed rumor detection SFCR method is illustrated in Fig. 1. It mainly consists of three parts: Comment Stance Learning, Non-Consecutive Semantic Features Learning, and Classification.

A. COMMENT STANCE LEARNING

Firstly, we utilize a pre-trained StanceBERTa model to obtain the stance information of each comment. StanceBERTa is a language model developed for stance detection, which involves identifying the author's attitude or viewpoint toward a specific target or topic. StanceBERTa is based on the BERT architecture, a widely used pre-trained language model for natural language processing tasks. Its performance has been proven to be particularly strong in detecting stances related

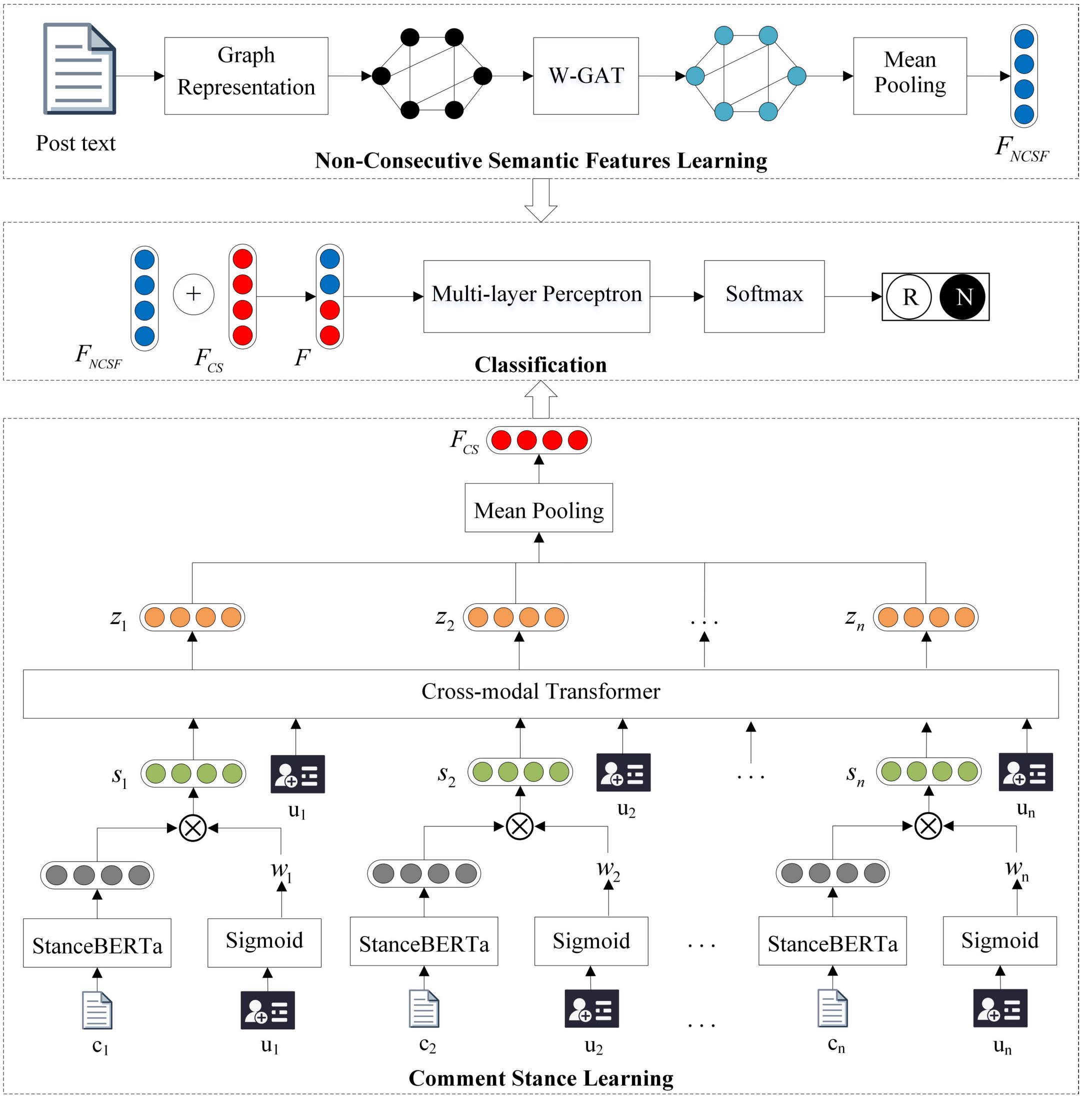


FIGURE 1. Overall framework of the SFCR.

to controversial topics, where detecting subtle nuances and differences in language is especially crucial.

Secondly, considering the stance of different users should be given different weights. For example, the weight of verified users with many followers should be higher than that of ordinary unverified users. We do not manually design the weight calculation but learn it automatically through neural networks. We use a fully connected layer to learn the weight values of each user and obtain the stance information s_i by combining the output of StanceBERTa and the weight values w_i.

$$w_i = Sigmoid(W_u \times u_i + b_u) \tag{1}$$

$$s_i = w_i \times StanceBERTa(c_i) \tag{2}$$

where W_u and b_u are the parameters and bias terms of the fully connected layer, u_i represents the user features of the i-th comment, and c_i represents the content information of the i-th comment.

Then, according to the chronological order of comments, we input the temporal data $S_{p_i} = \{s_1, s_2, \ldots, s_n\}$ of the comment stances on post p_i and the corresponding user sequence data $U_{p_i} = \{u_1, u_2, \ldots, u_n\}$ into the Cross-modal Transformer [37] to learn the temporal features $Z \in R^{n \times d_v}$ of stances fused with users' features.

$$Z = Crossmodal\_Transformer(S_{p_i}, U_{p_i}) \tag{3}$$

Cross-modal Transformer is an extension of the Transformer that can handle tasks with information from multiple modalities. We use it to process two modal information: comments and users. The architecture of the Cross-modal Transformer is illustrated in Fig.2.

The Cross-modal Transformer is a deep stacking of several cross-modal attention blocks. The structure of cross-modal attention is shown in Fig.3.

Where $X_\alpha \in R^{T_\alpha \times d_\alpha}$ and $X_\beta \in R^{T_\beta \times d_\beta}$ are the input sequences of modality α and modality β, respectively, T_α and

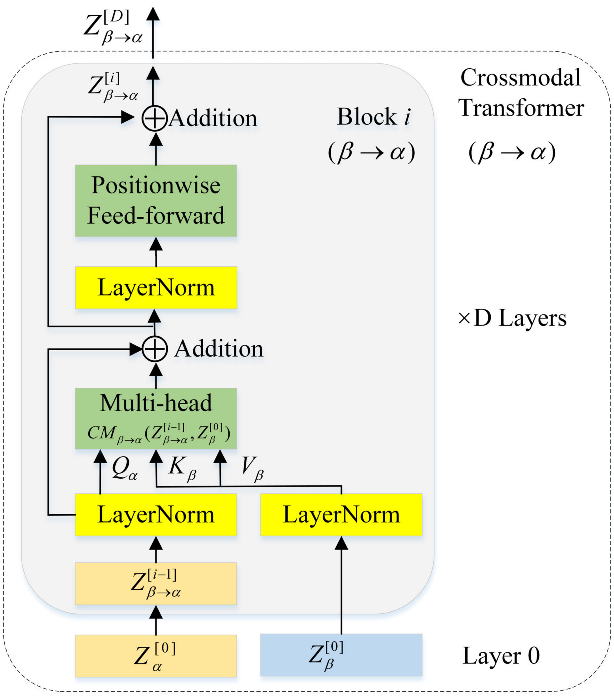


FIGURE 2. Overall architecture for crossmodal transformer.

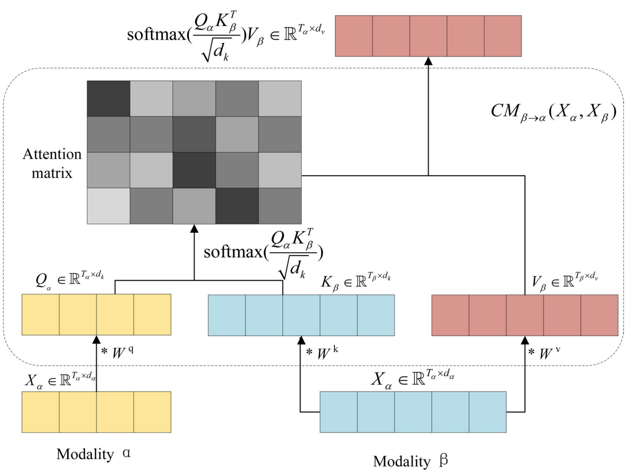


FIGURE 3. Cross-modal attention.

T_β are the lengths of the sequences, and d_α and d_β are the feature dimensions. $Q_\alpha = X_\alpha W^q$, $Q_\alpha \in R^{T_\alpha \times d_k}$ is Querys, $K_\beta = X_\beta W^k$, $K_\beta \in R^{T_\beta \times d_k}$ is Keys, and $V_\beta = X_\beta W^v$, $V_\beta \in R^{T_\beta \times d_v}$ is Values. Here $W^q \in R^{d_\alpha \times d_k}$, $W^k \in R^{d_\beta \times d_k}$, and $W^v \in R^{d_\beta \times d_v}$ are the weight parameters that need to be learned. Finally, we use global average pooling to obtain the final stance information $F_{CS} \in R^{d_v}$ for the comments.

$$F_{CS} = Mean\_Pooling(Z) \tag{4}$$

1) NON-CONSECUTIVE SEMANTIC FEATURES LEARNING

To learn the non-consecutive semantic features of the post's text, we first convert the post's text into a graph representation $G = (V, E)$, where taking words as nodes V, the correlation between words as edges E, and the degree of correlation between words as edge weights. We use pointwise mutual information (PMI) to calculate the degree of correlation between words. In detail, we utilize a fixed-size window to collect co-occurrence statistics of words in all posts, and the calculation of PMI for word pairs is shown in (5) - (7).

$$p(w_i) = \frac{|W(w_i)|}{|W|} \tag{5}$$

$$p(w_i, w_j) = \frac{\left|W(w_i, w_j)\right|}{|W|} \tag{6}$$

$$PMI(w_i, w_j) = log\frac{p(w_i, w_j)}{p(w_i)p(w_j)} \tag{7}$$

Here $|W|$ represents the total number of sliding windows, $|W(w_i)|$ represents the number of sliding windows including the word w_i, and $\left|W(w_i, w_j)\right|$ represents the number of sliding windows including both w_i and w_j simultaneously. We use statistics data based on the global corpus rather than a specific post's content. The PMI value reflects the correlation between words, where a positive PMI value indicates a high semantic correlation. Therefore, we only keep edges with positive PMI values and discard those with non-positive PMI values, as displayed in (8).

$$A_{i,j} = \begin{cases} PMI(w_i, w_j) & PMI(w_i, w_j) > 0 \\ 0 & PMI(w_i, w_j) \leqslant 0 \end{cases} \tag{8}$$

After obtaining the graph representation G and the adjacency weight matrix A of the post's text, we use a Weighted Graph Attention Network (W-GAT) [38], which considers edge weights, to learn non-consecutive semantic features. Unlike the standard GAT model, the W-GAT considers not only the feature similarity between nodes but also edge weights as an additional characteristic when calculating attention coefficients, thereby obtaining more comprehensive global information. The calculation of W-GAT is shown in (9).

$$h_i^{(l+1)} = \sigma(\sum_{j \in N(i)} a_{i,j}^l W^l h_j^l) \tag{9}$$

where σ represents the activation function, $N(i)$ represents the set of neighboring nodes of node v_i, W^l represents the learning parameters of the W-GAT at the l-th layer, h_j^l represents the output of neighboring nodes v_j in the previous layer, $a_{i,j}^l$ represents the weight of neighboring nodes v_j to node v_i in the l-th layer, which is calculated as shown in (10). We use Word2vec as the initial feature representation h_i^0 for nodes v_i.

$$a_{i,j}^l = softmax(e_{i,j}^l) \tag{10}$$

$$e_{i,j}^l = \alpha(h_i^l, h_j^l, A_{i,j}) \tag{11}$$

Here, $e_{i,j}^l$ represents the attention coefficient between node v_i and node v_j in the l-th layer of W-GAT, which is calculated as shown in (11). Here, α is a learnable function, and $A_{i,j}$ represents the weight of the edge linking nodes v_i and v_j,

TABLE 1. Datasets statistics.

Datasets	Rumors	Non-rumours	Total
Weibo	2313	2351	4664
PHEME	1972	3830	5802

which is the PMI value between words w_i and w_j. After L layers of W-GAT processing, we employ global average pooling to aggregate the feature $h_i^{(l+1)}$ of each node v_i in the graph to obtain the non-consecutive semantic features $F_{NCSF} \in R^{d^{(l+1)}}$, as calculated in (12). Here, V represents the set of nodes, and $|V|$ represents the size of V.

$$F_{NCSF} = \frac{1}{|V|} \sum_{i \in V} h_i^{(l+1)} \tag{12}$$

2) CLASSIFICATION

We concatenate the comment stance feature F_{CS} and non-consecutive semantic feature F_{NCSF} to obtain the final fused feature F, as shown in (13). Then, we input the fused feature F into the Multi-Layer Perceptron (MLP) and a soft-max layer to obtain the outputs, as shown in (14).

$$F = F_{CS} \oplus F_{NCSF} \tag{13}$$

$$\hat{y} = softmax(W \times MLP(F) + b) \tag{14}$$

where W and b are linear layer parameters and bias terms, respectively. We train the model by minimizing the cross-entropy loss, as shown in (15).

$$L = -\frac{1}{N} \sum_{i=1}^{N} (y_i log \hat{y}_i + (1 - y_i) log(1 - \hat{y}_i)) + \frac{\lambda}{2} \|W\|_2^2 \tag{15}$$

Here y_i is the real label value of the post p_i, $\hat{y}_i$ is the model's predicted value, $\frac{\lambda}{2} \|W\|_2^2$ is $L2$ regularization to reduce the degree of overfitting, and W is the model parameters.

IV. EXPERIMENTS

A. EXPERIMENTAL SETUP

1) EXPERIMENTAL DATA AND EVALUATION METRICS

The experiment uses two public real-world social media rumor datasets, Weibo[4] and PHEME[5], to validate the effectiveness of our proposed SFCR method. The statistical information of the PHEME and Weibo is shown in Table 1. We split the datasets into training, validation, and testing sets with a ratio of 3:1:1, and we use K-fold cross-validation to evaluate the performance of the methods. We use average accuracy, rumor precision, rumor recall, and F1-score as the performance evaluation metrics.

2) IMPLEMENTATION DETAILS

In the comment stance learning module, we choose verified status, number of friends, number of followers, registration time, and number of posts as the user features. The number of layers in the Cross-modal Transformer is set to 4, the number of attention heads is set to 2, and the hidden unit size of each layer is set to 64. In the non-consecutive semantic feature learning module, the sliding window size for calculating PMI for English and Chinese texts is set to 6 and 9, respectively. We use a 2-layer W-GAT, and the output dimensions of the first and second layers are 128 and 64, respectively. In the classification network, we use a 2-layer perceptron, with the output dimensions of the first and second layers set to 64 and 32, respectively. During model training, we use the Adam optimizer with a learning rate of 0.001, a batch size of 16, and 5-fold cross-validation. The main hyperparameter for the model and training are exhibited in Table 2.

TABLE 2. Hyperparameters setting.

Type	Name	Value
Model parameters	# of Cross-modal blocks	4
	# of Cross-modal attention heads	2
	Cross-modal transformer hidden unit size	64
	# of layers of the W-GAT	2
	Output dimension of the W-GAT	128,64
	Sliding window size for chinese text	9
	Sliding window size for english text	6
	Word2vec dimension	300
	# of layers MLP	2
	Output dimension of MLP	64,32
Training parameters	Learning rate	1e-3
	Regularization parameter	1e-5
	Maximum training epochs	30
	Early stopping patience	8
	Batch size	16
	Dropout rate	0.1
	K-fold cross-validation	5
	Optimizer	Adam

B. RESULTS AND ANALYSIS

1) COMPARATIVE ANALYSIS

To validate the effectiveness of our proposed SFCR method, we select the rumor detection methods in Table 3 as the baselines for comparison. The experimental results on the Weibo and PHEME datasets are displayed in Tables 4 and 5, respectively.

From the experimental results in Tables 4 and 5, we find that all the deep learning-based rumor detection methods outperform the statistical machine learning-based method SVM-TS, which depends on manual feature engineering. It is because manually designed features lack flexibility and comprehensiveness and cannot represent the deep semantic features of rumors. Deep learning methods automatically learn high-level abstract features fitting for tasks in a data-driven way, which has more robust adaptability and extensive diversity. The performance of rumor detection methods based on propagation features is generally better than those based on content features because the former can effectively reflect the process and path of rumor propagation, capture social relationships and interactions in social networks, and thus more accurately depict the characteristics of rumor propagation. However, the propagation feature-based methods need rumors to be spread to a certain extent to achieve good

[4]https://www.dropbox.com/s/46r50ctrfa0ur1o/rumdect.zip?dl=0

[5]https://figshare.com/articles/dataset/PHEME_dataset_of_rumours_and_non-rumours/4010619

TABLE 3. Baseline methods.

Methods	Years	Algorithms	Features Content	Features Propagation	Features User
SVM-TS [39]	2015	SVM	✓	✓	✓
GRU-2 [15]	2016	GRU		✓	
LSTM-CNN [3]	2018	LSTM+CNN	✓		
TDRD [6]	2020	CNN	✓		
LSTM-Attention [8]	2020	LSTM+Attention	✓		✓
GAN_based [5]	2021	GAN	✓		
EGCN [21]	2021	GCN	✓	✓	
User-Reply-GCN [32]	2021	GCN		✓	✓

TABLE 4. Experimental results on Weibo dataset.

Methods	Accuracy	Precision	Recall	F1-score
SVM-TS	84.4	86.0	85.3	85.6
LSTM-CNN	85.2	85.7	86.1	85.9
TDRD	86.7	87.0	86.0	86.5
GAN-based	86.8	86.5	89.4	87.9
LSTM-Attention	89.6	89.9	91.6	90.7
GRU-2	90.1	87.6	91.6	89.6
EGCN	91.3	89.9	92.0	90.9
User-Reply-GCN	91.6	90.2	92.3	91.2
SFCR	**93.4**	**92.3**	**93.8**	**93.0**

TABLE 5. Experimental results on PHEME dataset.

Methods	Accuracy	Precision	Recall	F1-score
SVM-TS	78.3	69.2	73.1	71.1
LSTM-CNN	80.4	80.1	81.1	80.6
TDRD	82.7	81.3	78.6	79.9
GAN-based	82.7	81.6	79.1	80.3
LSTM-Attention	83.0	82.3	81.6	81.9
GRU-2	82.6	82.1	81.0	81.5
EGCN	83.5	82.5	82.4	82.4
User-Reply-GCN	83.6	82.2	83.3	82.7
SFCR	**85.1**	**84.2**	**85.4**	**84.8**

performance, and early detection may be less effective. The performance of LSTM-Attention, which integrates both content and user features, is better than that of the three methods based solely on content features, namely LSTM-CNN, TDRD, and GAN-based, proving that user features play a crucial role in rumor detection. The performance of EGCN, which integrates content features into propagation features, and User-Reply-GCN, which integrates user features into propagation features, is better than that of GRU-2, based solely on propagation features, indicating the effectiveness of integrating content and user features into propagation features. Our proposed method, SFCR, integrates user, content, and propagation features, achieving the highest values on all metrics and demonstrating the effectiveness of our method.

2) ABLATION EXPERIMENTAL ANALYSIS

We conduct five ablation experiments to investigate the contribution of different modules to SFCR. Models 1 and 2 remove non-consecutive semantic feature F_{NCSF} and comment stance feature F_{CS}, respectively. Model 3 omits the comment user features in learning comment stance features. Model 4 replaces the non-consecutive semantic feature F_{NCSF} learned based on W-GAT with the consecutive semantic feature F_{CSF} extracted based on BERT. Model 5 replaces the W-GAT with a standard graph attention neural network. The decline in accuracy, precision, recall, and F1-score of each model compared to SFCR are displayed in Table 6.

We can observe from Table 6 that each module plays a unique role, and removing or replacing any of them would affect the SFCR's performance. The feature F_{NCSF} learned by the W-GAT can effectively represent the non-consecutive semantic features of the post's text, so removing it will significantly impact the method's performance. The feature F_{CS}, learned by StanceBERTa and Cross-modal Transformer, contains stance information from different users on the post, which plays a critical role in rumor detection. Therefore, removing it will significantly degrade the performance of the method. The features, such as whether the user is verified, their number of followers, friends, registration time, and post count, can reflect the user's authority. The attention mechanism can make our model focus more on the stance of authoritative users on the post, avoiding intentional hype from internet trolls. Therefore, not differentiating the stance weights of different users will also degrade the method's performance. After replacing the non-consecutive semantic feature F_{NCSF} with the consecutive semantic feature F_{CSF}, the method's performance decreases because social media posts often exhibit discrete and fragmented characteristics. The feature F_{NCSF} learned by graph neural networks can better represent the non-consecutive and long-distance dependent semantic features. When the W-GAT is replaced with GAT, the method's performance decreases because W-GAT considers feature similarity between nodes and takes the edge weight as an additional factor when calculating the attention coefficient, enabling it to obtain more accurate global information.

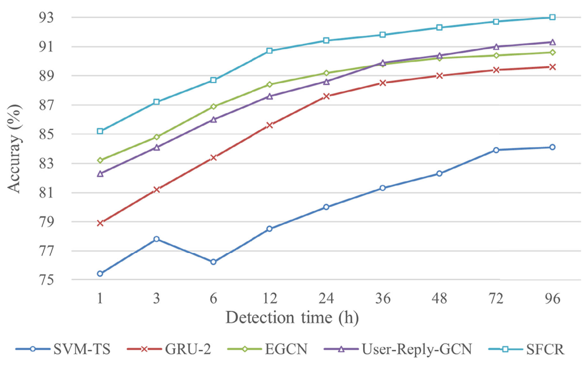
3) EARLY DETECTION ANALYSIS

To verify the performance of our proposed SFCR method in early rumor detection tasks, we select nine time points (1h, 3h, 6h, 12h, 24h, 36h, 48h, 72h, and 96h) after the post is posted as detection points. At each detection point, the comment data input to the model is before that time point. We select SVM-TS, GRU-2, EGCN, and User-Reply-GCN based on propagation features as baselines. The experimental results are displayed in Fig.4.

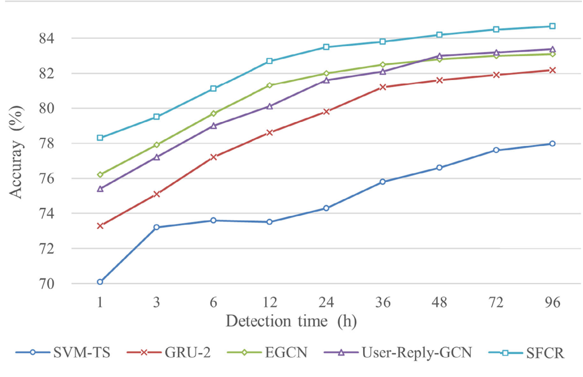
Over time, the accuracy of each method shows a varying degree of improvement. Compared to SVM-TS, the accuracy of GRU-2, EGCN, User-Reply-GCN, and SFCR increases more rapidly and steadily. After 36 hours, their accuracy

TABLE 6. Ablation Experiment Results.

Models	Weibo $\Delta_{accuracy}$	Weibo $\Delta_{precision}$	Weibo Δ_{recall}	Weibo $\Delta_{F1-score}$	PHEME $\Delta_{accuracy}$	PHEME $\Delta_{precision}$	PHEME Δ_{recall}	PHEME $\Delta_{F1-score}$
① (-)F_{NCSF}	-1.7	-1.2	-1.1	-1.2	-1.3	-1.4	-1.2	-1.3
② (-)F_{CS}	-2.1	-2.2	-2.2	-2.2	-1.8	-1.6	-1.4	-1.5
③ (-)$Users$	-1.2	-0.9	-0.8	-0.8	-1.0	-0.9	-0.7	-0.8
④ $F_{NCSF}\rightarrow F_{CSF}$	-0.8	-0.7	-0.6	-0.7	-0.7	-0.6	-0.6	-0.6
⑤ W-GAT→GAT	-0.6	-0.5	-0.5	-0.5	-0.5	-0.4	-0.6	-0.5



(a) Early detection results on Weibo dataset



(b) Early detection results on PHEME dataset

FIGURE 4. Early detection results.

approaches the maximum value, while SVM-TS requires 76 hours, demonstrating the advantages of deep learning methods over feature engineering methods in early rumor detection. In the early detection within 1-24 hours, the accuracy of EGCN, User-Reply-GCN, and SFCR is significantly higher than that of GRU-2. Because the former three methods integrate content and user features, they can reduce the dependence on propagation features, thereby improving early detection performance. SFCR shows higher accuracy than the other four models at all time points, achieving good detection results within 24 hours of the rumor being released, while the other methods require more than 36 hours. Our method detects the rumor at least 12 hours earlier, proving its effectiveness in early rumor detection tasks.

V. CONCLUSION

Social media has facilitated communication among people while fueling the proliferation and dissemination of online rumors. Automatic rumor detection models are paramount in mitigating the spread of rumors. Currently, most deep learning rumor detection methods based on semantic features fail to consider social media posts' discrete and fragmented nature. Meanwhile, time series models based on propagation features do not differentiate between the importance of various commenting users. Therefore, we propose a novel rumor detection method that integrates non-consecutive semantic features and stances considering user weights. Non-consecutive semantic features of posts are learned through a weighted graph attention network, while stance features of comments that incorporate user weights are learned using StanceBERTa and Cross-modal Transformer. The experimental results on two public social media rumor datasets indicate that our method outperforms the baseline methods. Specifically, our method detects rumors at least 12 hours earlier than the baseline methods, demonstrating its efficacy in early rumor detection. While our method integrates multiple features, it does not consider the image or video information that may be attached to the post. Therefore, our subsequent research will explore how to achieve multi-modal rumor detection.

REFERENCES

[1] A. M. Almars, M. Almaliki, T. H. Noor, M. M. Alwateer, and E. Atlam, "HANN: Hybrid attention neural network for detecting COVID-19 related rumors," *IEEE Access*, vol. 10, pp. 12334–12344, 2022.

[2] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet—A deep convolutional neural network for fake news detection," *Cogn. Syst. Res.*, vol. 61, pp. 32–44, Jun. 2020.

[3] O. Ajao, D. Bhowmik, and S. Zargari, "Fake news identification on Twitter with hybrid CNN and RNN models," in *Proc. 9th Int. Conf. Social Media Soc.*, Copenhagen, Denmark, Jul. 2018, pp. 226–230.

[4] S. A. Alkhodair, S. H. H. Ding, B. C. M. Fung, and J. Liu, "Detecting breaking news rumors of emerging topics in social media," *Inf. Process. Manage.*, vol. 57, no. 2, Mar. 2020, Art. no. 102018.

[5] M. Cheng, Y. Li, S. Nazarian, and P. Bogdan, "From rumor to genetic mutation detection with explanations: A GAN approach," *Sci. Rep.*, vol. 11, no. 1, p. 5861, Mar. 2021.

[6] F. Xu, V. S. Sheng, and M. Wang, "Near real-time topic-driven rumor detection in source microblogs," *Knowledge-Based Syst.*, vol. 207, Nov. 2020, Art. no. 106391.

[7] T. Ma, H. Zhou, Y. Tian, and N. Al-Nabhan, "A novel rumor detection algorithm based on entity recognition, sentence reconfiguration, and ordinary differential equation network," *Neurocomputing*, vol. 447, pp. 224–234, Aug. 2021.

[8] J. P. Singh, A. Kumar, N. P. Rana, and Y. K. Dwivedi, "Attention-based LSTM network for rumor veracity estimation of tweets," *Inf. Syst. Frontiers*, vol. 24, no. 2, pp. 459–474, Aug. 2020.

[9] R. K. Kaliyar, A. Goswami, and P. Narang, "DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network," *J. Supercomput.*, vol. 77, no. 2, pp. 1015–1037, Feb. 2021.

[10] Y. Shao, J. Sun, T. Zhang, Y. Jiang, J. Ma, and J. Li, "Fake news detection based on multi-modal classifier ensemble," in *Proc. MAD*, New York, NY, USA, 2022, pp. 78–86.

[11] X. Chen, F. Zhou, G. Trajcevski, and M. Bonsangue, "Multi-view learning with distinguishable feature fusion for rumor detection," *Knowledge-Based Syst.*, vol. 240, Mar. 2022, Art. no. 108085.
[12] M. Davoudi, M. R. Moosavi, and M. H. Sadreddini, "DSS: A hybrid deep model for fake news detection using propagation tree and stance network," *Exp. Syst. Appl.*, vol. 198, Jul. 2022, Art. no. 116635.
[13] P. Zhang, H. Ran, C. Jia, X. Li, and X. Han, "A lightweight propagation path aggregating network with neural topic model for rumor detection," *Neurocomputing*, vol. 458, pp. 468–477, Oct. 2021.
[14] J. Ma, W. Gao, and K. F. Wong, "Detect rumors in microblog posts using propagation structure via kernel learning," in *Proc. ACL*, Vancouver, BC, Canada, 2017, pp. 717–780.
[15] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K. F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," in *Proc. AAAI*, New York, NY, USA, 2016, pp. 3818–3824.
[16] Y. Xu, C. Wang, Z. Dan, S. Sun, and F. Dong, "Deep recurrent neural network and data filtering for rumor detection on sina Weibo," *Symmetry*, vol. 11, no. 11, p. 1408, Nov. 2019.
[17] J. Ma, W. Gao, and K. F. Wong, "Rumor detection on Twitter with tree-structured recursive neural networks," in *Proc. ACL*, Melbourne, VIC, Australia, 2018, pp. 1980–1989.
[18] J. Ma, W. Gao, S. Joty, and K.-F. Wong, "An attention-based rumor detection model with tree-structured recursive neural networks," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 4, pp. 1–28, Aug. 2020.
[19] T. Chen, X. Li, H. Yin, and J. Zhang, "Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection," in *Proc. PAKDD*, Melbourne, VIC, Australia, 2018, pp. 40–52.
[20] Q. Huang, C. Zhou, J. Wu, L. Liu, and B. Wang, "Deep spatial–temporal structure learning for rumor detection on Twitter," *Neural Comput. Appl.*, vol. 35, no. 18, pp. 12995–13005, Jun. 2023.
[21] N. Bai, F. Meng, X. Rui, and Z. Wang, "Rumour detection based on graph convolutional neural net," *IEEE Access*, vol. 9, pp. 21686–21693, 2021.
[22] Z. Wu, D. Pi, J. Chen, M. Xie, and J. Cao, "Rumor detection based on propagation graph neural network with attention mechanism," *Exp. Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113595.
[23] J. Choi, T. Ko, Y. Choi, H. Byun, and C.-K. Kim, "Dynamic graph convolutional networks with attention mechanism for rumor detection on social media," *PLoS ONE*, vol. 16, no. 8, Aug. 2021, Art. no. e0256039.
[24] H. F. Alsaif and H. D. Aldossari, "Review of stance detection for rumor verification in social media," *Eng. Appl. Artif. Intell.*, vol. 119, Mar. 2023, Art. no. 105801.
[25] N. Bai, Z. Wang, and F. Meng, "A stochastic attention CNN model for rumor stance classification," *IEEE Access*, vol. 8, pp. 80771–80778, 2020.
[26] J. Ma, W. Gao, and K. F. Wong, "Detect rumor and stance jointly by neural multi-task learning," in *Proc. Web Conf.*, Lyon, France, 2018, pp. 585–593.
[27] C. Li, H. Peng, J. Li, L. Sun, L. Lyu, L. Wang, P. S. Yu, and L. He, "Joint stance and rumor detection in hierarchical heterogeneous graph," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2530–2542, Jun. 2022.
[28] K. Ye, Y. Piao, K. Zhao, and X. Cui, "Graph enhanced BERT for stance-aware rumor verification on social media," in *Proc. ICANN*, Bratislava, Slovakia, 2021, pp. 422–435.
[29] X. Han, Z. Huang, M. Lu, D. Li, and J. Qiu, "Rumor verification on social media with stance-aware recursive tree," in *Proc. KSEM*, Tokyo, Japan, 2021, pp. 149–161.
[30] A. Khandelwal, "Fine-tune longformer for jointly predicting rumor stance and veracity," in *Proc. 3rd ACM India Joint Int. Conf. Data Sci. Manage. Data (8th ACM IKDD CODS 26th COMAD)*, Jan. 2021, pp. 10–19.
[31] 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," *Inf. Sci.*, vol. 560, pp. 137–151, Jun. 2021.
[32] S. Lotfi, M. Mirzarezaee, M. Hosseinzadeh, and V. Seydi, "Detection of rumor conversations in Twitter using graph convolutional networks," *Appl. Intell.*, vol. 51, no. 7, pp. 4774–4787, Jan. 2021.
[33] M. Sun, X. Zhang, J. Zheng, and G. Ma, "DDGCN: Dual dynamic graph convolutional networks for rumor detection on social media," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 4, Jun. 2022, pp. 4611–4619.
[34] S. Xu, X. Liu, K. Ma, F. Dong, B. Riskhan, S. Xiang, and C. Bing, "Rumor detection on social media using hierarchically aggregated feature via graph neural networks," *Int. J. Speech Technol.*, vol. 53, no. 3, pp. 3136–3149, Feb. 2023.
[35] W. Cui and M. Shang, "KAGN: Knowledge-powered attention and graph convolutional networks for social media rumor detection," *J. Big Data*, vol. 10, no. 1, p. 45, Apr. 2023.
[36] P. Zheng, Z. Huang, Y. Dou, and Y. Yan, "Rumor detection on social media through mining the social circles with high homogeneity," *Inf. Sci.*, vol. 642, Sep. 2023, Art. no. 119083, doi: 10.1016/J.INS.2023.119083.
[37] Y. H. H. Tsai, "Multimodal transformer for unaligned multimodal language sequences," in *Proc. ACL*, Florence, Italy, 2019, pp. 6558–6569.
[38] E. Inan, "ZoKa: A fake news detection method using edge-weighted graph attention network with transfer models," *Neural Comput. Appl.*, vol. 34, no. 14, pp. 11669–11677, Mar. 2022.
[39] J. Ma, W. Gao, Z. Wei, Y. Lu, and K. F. Wong, "Detect rumors using time series of social context information on microblogging," in *Proc. CIKM*, Melbourne, VIC, Australia, 2015, pp. 1751–1754.

YI ZHU received the B.S. degree in auditing from the Pearl River College, Tianjin University of Finance and Economics, Tianjin, China, in 2022. He is currently pursuing the master's degree in administrative management with the Jiangxi University of Finance and Economics. His research interests include public crisis management and internet public opinion analysis.

GENSHENG WANG received the M.S. degree in computer software and theory from Jiangxi Normal University, Nanchang, China, in 2005, and the Ph.D. degree in management science and engineering from the Jiangxi University of Finance and Economics, Nanchang, in 2011.

From 2014 to 2021, he was an Assistant Professor and a Master Tutor of Jiangxi Finance and Economics. Since 2021, he has been a Professor with the Jiangxi University of Finance and Economics. He has published more than 20 research articles on internet public opinion and presided more than two National Natural Science Foundation of China projects. His current research interests include internet public opinion and data mining.

SHENG LI received the B.S. and M.S. degrees in administrative management and management science and engineering from Hunan University, Hunan, China, in 2007 and 2011, respectively.

Since 2017, he has been an Associate Professor with the Jiangxi University of Finance and Economics. In 2021, he was a Doctoral Supervisor with the Jiangxi University of Finance and Economics. He has published seven research articles on emergency management, two monographs, and presided over one National Social Science Fund Project and one National Social Science Fund major project sub-project. His current research interests include emergency management, environmental governance, and social governance.

XUEJIAN HUANG received the B.S. and M.S. degrees in information security and software engineering from Nanchang University, Nanchang, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Nanjing University of Information Science and Technology, Nanjing, China.

He is a Lecturer with the Computer Department, Jiangxi University of Finance and Economics, Nanchang. His current research interests include NLP and data mining, especially in sentiment and social network analysis tasks.