RESEARCH

Open Access

RPf-GCNs: reciprocal perspective driven fused GCNs for rumor detection on social media

Zafran Khan[1], Jeonghwan Gwak[2], Naima Iltaf[3], Witold Pedrycz[4,5,6] and Moongu Jeon[1*]

*Correspondence: mgjeon@gist.ac.kr

[1] School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea
[2] Department of Software, Korea National University of Transportation, Chungju 27469, South Korea
[3] Department of Computer Software Engineering, National University of Sciences and Technology, Islamabad, Pakistan
[4] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada
[5] Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Sariyer, Istanbul, Turkey
[6] Systems Research Institute, Polish Academy of Sciences, Istinye University, 00-901 Warsaw, Poland

Abstract

The earliest detection of rumors across social media is the need to the hour in present global village. User's are seamlessly connected in an unstructured network leading to rapid flow of information. User's on the social media with malign intents may share defamatory content to contribute towards the fifth generation media warfare. The ingress of such defamatory content into society can result in panic, uncertainty and demoralization the peoples. Due to the huge amount of content over social platforms, the detection of malicious contents is hard. Earlier research while focuses on content profiling and flow of information, however, the reciprocal perspective of the source and following contents is missing. In this research, a novel Reciprocal Perspective fused Graph Convolutional Neural Network (RPf-GCN) is proposed. The proposed framework incorporates twin GCNs to encode both the bottom-up and top-down perspectives, enhancing the understanding of rumor propagation. Moreover convolutional operation is employed to fuse reciprocal perspective, providing a holistic view of the conversations. To validate the efficacy of the proposed framework, we conducted a series of experiments using real-world datasets, including PHEME and SemEval. Experimentation performed illustrates that the proposed framework outperformed over various baselines in two different evaluation metrics namely Macro F1 (for PHEME 0.736, for SemEval 0.461) and Accuracy (for PHEME 0.748, for SemEval 0.658).

Keywords: Social media analytic, Rumor detection, Conversation reciprocal perspective, Graph convolutional network

Introduction

Contents available on social media platforms can effectively exploit the opinions of readers. Such content can be cleverly manipulated to spread propaganda, rumors, and other misinformation. The spread of such propagandist content in society can lead to fear, uncertainty, panic, or even financial losses in trading markets. Psychological research shows that human beings are only 55–58% capable of identifying malicious content [1], which is a clear indicator of how easily the public can be deceived. Detection of such content can protect society from being jeopardized by such misinformation campaigns. Rumors on social media can be classified as true, false, or unverified based on the authenticity of the facts presented [2–5]. Social media platforms lack an effective verification mechanism for the content shared by users, allowing them to spread defamatory

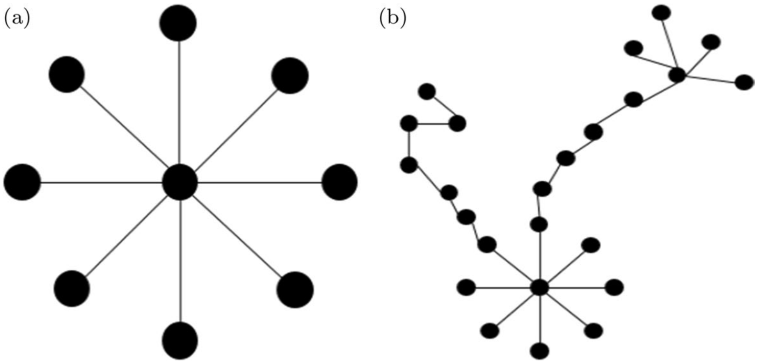


Fig. 1 Propagation pattern of **a**: False rumor, **b**: True rumor

content without hindrance. Thus, devising an automated framework for their earliest detection is the dire need of the hour in this digital era.

Usually it happens that users start commenting on a post made by a source user. Their responses, in the form of comments, show their consent, emotions, and viewpoints. Such comments or retweeted posts lead to conversation threads of different lengths, depending on the users involved in the comments. False (true) rumors mean that veracity of claim is false (true) [4]. Any conversation thread is based on the root node (source post) and the threads (comments) linked to the root node. It has been observed through dataset that shallow propagation is observed by false rumors, whereas true rumors show longer multi-branch and multi-point propagation, as shown in Fig. 1. Rumors need to be attractive enough to grab people's attention; thus, they are more likely to break out at the roots. On the other hand, true rumors have no intention of spreading, thus they have a scattered pattern of spreading.

Existing researchers have analyzed the sequential, structural, and temporal aspects of conversation threads assuming that such threads are of tree-structured or non-directed graphs. However, the direction of conversation threads is often ignored. It is to be noted that direction of conversations carry patterns of rumor content flow and comprehensiveness. The source node and follower nodes have a relation from different views, i.e., top-down and bottom-up. The top-down perspective involves rumors flowing from the source post to followers' comments, while the bottom-up perspective involves the opposite. Therefore, it is necessary to exploit both views to synchronously encode both top-down and bottom-up flows.

This research proposed a reciprocal perspective aware graph convolutional neural network (GCN). The inspiration for this proposed framework is traced back to the field of computer vision, where a colored image is a seamless fusion of its three channels, with each channel representing an individual perspective of the image. The analogy would be such that each of the channel is the individual perspective of the image. The proposed framework consists of twin GCN that encodes both the bottom-up and top-down perspectives. To effectively fuse these views, a convolutional operation is employed to capture the reciprocal perspective. The key contribution of this research is proposing a reciprocal perspective-driven GCN that effectively learns and fuses the reciprocal perspectives of conversation. Moreover, a series of experiments performed on two real-world datasets, i.e., PHEME and SemEval, aimed to prove the efficacy of the proposed framework.

The paper is structured with sections covering literature review (Sect. Literature review), formulation of problem statement (Sect. Problem statement formulation), detailed methodology (Sect. Proposed framewrok), experiments/results (Sect. results), conclusions (Sect. Conclusion). At last Sect. Conclusion concludes the paper followed by suggested future work.

Literature review

Online social media (OSM) is an ocean of information in the form of users and the content shared by them. The presence of fake news, rumors, and propagandist content on OSM platforms is no surprise in today's era of global digitization. Thus, detecting such malicious content is considered prudent to safeguard our society against the possible spread of panic, fear, or even economic loss. For years, academic researchers have been focusing on this domain, and ample of research work can be found in the literature.

Conventional techniques for identifying rumors typically involve extracting features from text, user profiles, and retweet propagation [31–36]. Ma et al. [13] utilized time series models to capture social context changes and kernel methods to create tree structures to represent propagation patterns. Nonetheless, these methods require significant feature engineering, which is both laborious and constrained.

In order to detect rumors or propaganda content, researchers have explored various domains with the aim of testing the performance of their frameworks. Ozbay and Alatas [8] implemented fake news detection in two steps. Initially they converted the unstructured data to a structured format and then applied various supervised learning algorithms by text mining methods. Kaliyar et al. [9] proposed a convolutional neural network based framework that automatically differentiate among the rumorous and normal contents on social media. The authors proposed deep neural network based approach that detects the fake news based on the shared contents, their context and related temporal information. Faustini et al. [10] suggested a framework that is independent of platform source and language for propaganda detection. It is entirely based on text features extraction techniques. The proposed framework was examined on three different language groups with optimal results. Four learning based algorithms namely random forest, KNN, SVM and Naive Bayes were implemented. Optimum results were obtained by random forest and SVM. Wang et al. [11] proposed a GNN model for early detection of fake news that is based on enhanced textual contents representation. The representation is achieved by integration of semantic relation and sequential ordering of textual contents. Liu and Wu [12] implemented deep learning based feature extraction, CNN classifier and a learning framework to detect the fake new at an early stages aimed to stop it spreading in the social media. They used datasets that were generated from twitter and Weibo.

The use of deep neural networks for rumor detection has been investigated by researchers. In particular, Ma et al. [14] used RNN to capture the sequential representations of textual posts at time interval. Liu et al. [15] combined RNN and CNN to extract user profile feature and deduce veracity of posts. Lu et al. [16] suggested the hybrid model that incorporates user profiles and source tweets. Yu et al. [24]

employed a hierarchical transformer framework to learn local and global interactions among shorter subthreads of longer conversation threads. However, these methods do not take into account the structural characteristics of conversation theams depicted in Fig. 1, that provide insights into how posts spread on social-media.

Ma et al. [4] introduced a model based on Recursive Neural Network (RvNN) that employs deep learning techniques to capture significant patterns from textual content and propagation structures. The model acquires latent representations of tweets within propagation trees through learning. Likewise, in their work, Lin et al. [25] employed undirected graph neural networks alongside multiple attention mechanisms to improve the learning of representations for individual posts and their interactions. However, considering the diverse relationships between nodes, the methods used encoded tree-structured graphs with only a single edge type. Consequently, in order to address this concern, Bian et al. [6] directed their attention towards both the top-down and bottom-up propagation relationships among nodes. Building upon Bian's work, Wei et al. [7] made further enhancements by eliminating unreliable relationships between nodes within rumor conversation threads. However, despite these advancements, these methods still face a limitation in effectively integrating multiple reciprocal views within rumor conversations to distinguish between false and true rumors from a global perspective.

Graph Neural Networks have gained popularity in recent years due to their ability to learn representations of structured data with high performance. They have been applied in various tasks, such as text classification [26] and recommendation systems [27]. Representative examples of graph neural networks include GCN [28] and GAT [29]. Authors in [37] utilized GCNs to highlight the ubiquitous presence of social circles in online social networks and their potential to reveal users' behavioral preferences. Drawing upon insights from information diffusion studies, substantial impact of social circles on the dynamics of rumor propagation, including its speed, reach, and content has been explored. Lin et al. [38] introduces a groundbreaking zero-shot framework, to identify rumors spanning diverse domains and languages. The approach begins by representing social media rumors as a collection of diverse propagation threads. Using GCN it incorporates domain-invariant structural features extracted from the propagation threads. This inclusion involves capturing structural position representations within influential community responses. The article [39] introduces a novel rumor detection model named "graph contrastive learning with feature augmentation" (FAGCL). This model aims to enhance rumor detection by introducing noise into the feature space and facilitating contrastive learning through the construction of asymmetric structures. FAGCL starts by using user preferences and news embeddings as the initial features of the rumor propagation tree. It then employs a graph attention network to iteratively update node representations. Sun et al. [40] introduced a novel approach called the "Knowledge-guided Dual-consistency Network." to detect rumors that incorporate multimedia content and focuses on capturing inconsistencies at two distinct levels: the cross-modal level and the content-knowledge level. It enables the robust learning of multi-modal representations, even when visual modality information is missing. To facilitate this, a unique token is introduced to differentiate between posts that contain visual modality and those that do not.

These methods are designed for single-view network representation, whereas in reality, there exist multi or reciprocal view networks, where each view corresponds to a different perspective of conversation thread. Consequently, considerable research efforts [30] have been devoted to the exploration of multi-view graph learning, with a specific focus on integrating node representations from each view into a comprehensive global node representation. In contrast, the current study investigates the fusion of features from reciprocal perspective graphs into a unified graph feature representation vector, aiming to detect rumors.

Problem statement formulation

The rumor detection task can be as follows. The social media is full of conversation threads that can be represented as $T = [t_1, t_2, t_3, \cdots, t_i, \cdots, t_p]$ where t_i represents the i^{th} conversation thread and p is the total number of threads existing in the dataset. Each t_i is composed of a source post s_i and various responses r_i; s.t $0 < i < n_i - 1$. Thus the overall structure of any i^{th} thread can be structured as $T_i = [s_1 : r_1, r_2, \cdots, r_{n_i-1}, G_i]$ .Here the term $G_i = \{N_i, E_i\}$ is a tree structure that is formed by the source post and responses, wherein $N_i = \{s_1 : r_1, r_2, \cdots, r_{n_i-1}\}$ are the number of nodes and $E_i = \{e_i \rightarrow e_j; 0 < i < n_i - 1, 0 < j < n_i - 1\}$ are the number of edges connecting the nodes e_i and e_j which are part of the conversation thread. In this article, the rumor detection problem is tackled as supervised learned mapping function $f = T_i \rightarrow L$ where T_i is a conversation thread having label L.

Proposed framework

The proposed framework is based upon reciprocal perspective-driven GCNs that classify the source post as either "rumorous" or not. Graph Convolutional Neural Networks (GCNs) have emerged as a powerful tool for analyzing structured data represented in the form of graphs or networks. Initially developed for tasks related to semi-supervised node classification and link prediction, GCNs have found application in various domains, including social network analysis, recommendation systems, and, more recently, rumor detection in social media.

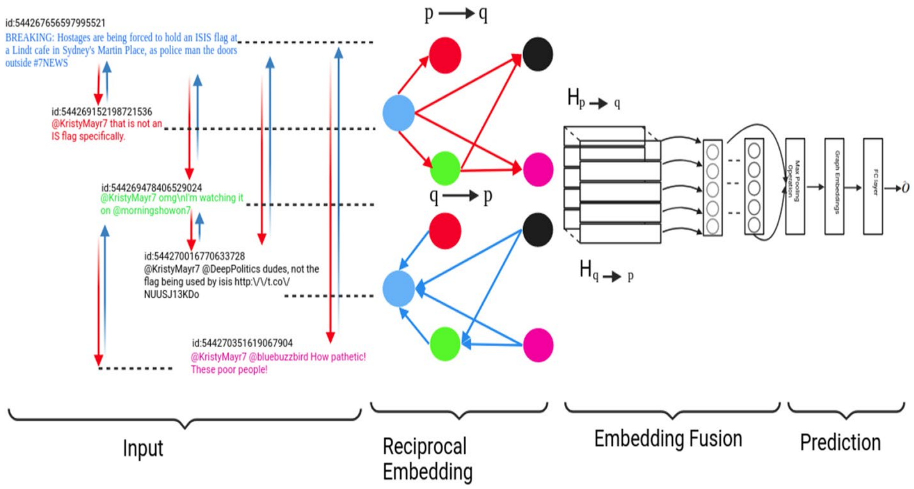
GCNs build upon the concept of convolutional neural networks, originally designed for regular grids such as images, and extend it to irregular data structures like graphs. The core idea is to learn node representations by aggregating information from neighboring nodes, enabling the network to capture complex relationships within the graph.

In a typical GCN, a node's representation is updated by combining features from its neighbors. This process can be mathematically defined through a propagation rule that considers both the node's own features and its adjacent nodes' features. The depth of these layers or iterations of propagation can be adjusted to control the model's capacity and ability to capture higher-order dependencies. The success of GCNs in graph-related tasks stems from their capacity to capture local and global information efficiently. This is especially relevant when analyzing social media data, where conversations and information propagation occur in a complex network structure.

An overview of the proposed framework is depicted in Fig. 2. It learns features from the reciprocal view of both the source and response posts. The proposed model undertakes three sequential tasks that involve reciprocal perspective embedding generation, fusion, and content classification. We will provide directions on how to apply the proposed framework to determine the veracity extent of a source post s_i in a conversation thread T_i. For the sake of simplicity, the subscript i will be removed in the following paragraphs.

Taking into consideration the graph $G = \{N, E\}$ of any conversation thread, the graphs can be represented by its adjacency matrix $\mathbf{A}_{p\rightarrow q} \in \boldsymbol{R}^{n\times n}$; $\mathbf{A}_{p\rightarrow q} = 1$ if node p has responded or commented to post of node q. Here the point to ponder is that though the adjacency matrix $\mathbf{A}_{p\rightarrow q}$ reflects the view of node p only who has responded to node q and leaves out the stance of node q towards node p. The proposed methodology attempts to fill up this gap by including the reciprocal perspective of both nodes for each other. The inclusion of the reciprocal perspective in our model is motivated by the need for a more comprehensive understanding of social media rumors. By analyzing conversations from both the source and follower viewpoints, we gain a holistic view of how rumors propagate and evolve. This perspective-driven approach allows us to capture a broader range of indicators and enhances the model's performance in detecting malicious content. It provides a more robust and nuanced analysis of rumor dynamics, making it a valuable addition to proposed framework.

Thus the adjacency matrix $\mathbf{A}_{p\rightarrow q}$ can be assumed to have the descending perspective of the graph from parent to child nodes. Similarly another adjacency matrix $\mathbf{A}_{q\rightarrow p}$ can also be constructed using the same graph $G = \{N, E\}$ wherein adjacency matrix $\mathbf{A}_{q\rightarrow p}$ would have the ascending perspective of the same graph from child to parent nodes. we have employed the concept of dropping some random percentile of edge [20] on G as of earlier researchers aimed to avoid overfitting of the proposed model. Both the adjacency matrices of descending perspective $\mathbf{A}_{p\rightarrow q}$ and ascending perspective $\mathbf{A}_{q\rightarrow p}$ share the same features matrix $\mathbf{F}$ for each node. The feature matrix for each node is prepared by using top-3000 words embedding generated by bi-LSTM. We use bi-LSTM because it generates the textual embedding



id:544267656597995521
BREAKING: Hostages are being forced to hold an ISIS flag at a Lindt cafe in Sydney's Martin Place, as police man the doors outside #7NEWS

id:544269152198721536
@KristyMayr7 that is not an IS flag specifically.

id:544269478406529024
@KristyMayr7 omg\nI'm watching it on @morningshowon7

id:544270016770633728
@KristyMayr7 @DeepPolitics dudes, not the flag being used by isis http:\/\/t.co\/NUUSJ13KDo

id:544270351619067904
@KristyMayr7 @bluebuzzbird How pathetic! These poor people!

Fig. 2 Proposed framework

while considering the text in both directions. Subsequently, the node embeddings are updated using two concurrent graph convolutional neural networks (GCNs). GCN's convolutional operations in each layer can be represented by Eqs. 1 and 2.

$$H_{p\to q}^{(l+1)} = \sigma\left(\mathbf{A}_{p\to q}\mathbf{F}W^{(l)}\right) \tag{1}$$

$$H_{p\to q}^{(l+2)} = \sigma\left(\mathbf{A}_{p\to q}H_{p\to q}^{(l+1)}W^{(l+1)}\right) \tag{2}$$

Here σ replicates the non-linear activation function typically it is ReLu, $\mathbf{A}_{p\to q}$ is the Adjacency matrix of $(n \times n)$, $\mathbf{F}$ is the feature matrix, $H_{p\to q}^{(l+1)} \in \boldsymbol{R}^{n\times D_1}$ and $H_{p\to q}^{(l+2)} \in \boldsymbol{R}^{n\times D_2}$ are the hidden features representations at layers $l+1$ and $l+2$ respectively. Lastly, $W^{(l)} \in \boldsymbol{R}^{D_1\times D_2}$ and $W^{(l+1)} \in \boldsymbol{R}^{D_1\times D_2}$ are the trainable weight matrices of layers l and $l+1$ respectively. Similar to descending perspective, the ascending perspective calculations can also be replicated by Eqs. 1 and 2 which would result in $H_{q\to p}^{(l+1)} \in \boldsymbol{R}^{n\times D_1}$ and $H_{q\to p}^{(l+2)} \in \boldsymbol{R}^{n\times D_2}$.

After obtaining these reciprocal perspectives of the conversation threads for each node, we combine them for further processing. As discussed earlier, the proposed framework is inspired by computer vision, and we consider these reciprocal perspectives as two channels of an RGB image, with each node corresponding to a pixel. Therefore, the feature representation of conversation threads can be expressed in the form:

$$H = \begin{bmatrix} H_{p\to q}^{(l+1)}, H_{p\to q}^{(l+2)}, \cdots, H_{p\to q}^{(n)} \\ H_{q\to p}^{(l+1)}, H_{q\to p}^{(l+2)}, \cdots, H_{q\to p}^{(n)} \end{bmatrix} \in \boldsymbol{R}^{2n\times D_2} \tag{3}$$

Influenced by the convolution operations of image channels in computer vision, H in (3) can be considered as bi-channel input to a convolution operation, that involves various filters $\boldsymbol{W}_f \in \boldsymbol{R}^{2l\times D_2}$ s.t $f \in [1, F]$. The convolution operation is applied to a window of y nodes which leads to generation of feature maps in accordance of the following operations as shown below:

$$\partial_i^f = \sigma(\boldsymbol{W}_f \otimes H_{i:i+y-1} + b) \tag{4}$$

Where $\boldsymbol{W}_f$ and b are optimized parameters, $\otimes$ indicates the bi-channels convolutional operations and σ is the ReLU activation function applied to the output.Thereafter, following feature map is obtained:

$$\partial^f = [\partial_1^f, \partial_2^f, \cdots, \partial_{n-y+1}^f] \tag{5}$$

Max-over pooling layer operation, max (∂^f), is applied to capture the most important feature, i.e., the highest value of ∂^f amongst the feature map i.e $\bar{\partial}^f = \max(\partial^f)$. The maximum feature values produced by all filters are concontinated and final representation of the conversation thread is obtained as shown in Eq. 6, where F is the total number of filters applied.

$$\boldsymbol{F} = [\bar{\partial}^f{}_1, \bar{\partial}^f{}_2, \cdots, \bar{\partial}^f{}_F] \tag{6}$$

After obtaining the final feature representation of the conversation thread, the feature map is fed into a fully connected layer with a softmax activation function. The purpose of this layer is to predict the probabilistic values whether the source post is a rumor or not. The functionality is represented as:

$$\hat{\boldsymbol{O}} = softmax(FC(\boldsymbol{F})) \tag{7}$$

The aim of the loss function being used during the training of the proposed framework is to minimize the cross-entropy among predicted and ground truth values.

$$\mathcal{L} = -\sum_{i}^{|\boldsymbol{O}|} \boldsymbol{O}^i \log \hat{\boldsymbol{O}}^i \tag{8}$$

Where $\boldsymbol{O}^i$, is the feature representation of source post s_i present in conversation thread T_i.

Experimentation results

This section elaborates on the experimental setup and results obtained from benchmark datasets, namely PHEME and SemEval. It also includes a comparative analysis of the proposed framework against its baselines. Furthermore, we conducted various ablation studies exploring the effects of different factors on the proposed framework.

Experimentation details

The proposed framework has been implemented on a desktop system installed with Ubuntu 18.04 LTS (Bionic), which has 16 GB of RAM, an AMD® Ryzen 7 3700x 8-core processor, and an NVIDIA GeForce RTX 3080 Ti GPU. During the training process, we set the dropout rate to 0.5, **F** to 64, and D_1 and D_2 to 64 as well. Additionally, the learning rate was set to 1×10^{-4}, and the batch size was set to 64. The framework was trained for 100 epochs with L2-regularization and a weight penalty of 0.001. The Adam optimizer was used during the training process.

Dataset description

In order to examine the effectiveness and validate the performance of the proposed model, two benchmark datasets were selected. The main reason for selecting these datasets is that they are versatile in nature and contain the requisite details of users and responses to source posts. Such details are primarily required to create conversation threads with a reciprocal effect between source and responses. The statistics of both benchmark datasets, post-preprocessing, are given in Table 1.

PHEME is a dataset based on rumors and non-rumors, consisting of nine real-time incidents that occurred between 2012 and 2015. The original incidents are comprised of tweets from a source user, to which various followers responded. The tweets are provided in a JSON file with 19 features corresponding to each tweet. In order to avoid over-fitting and ensure the convergence of our proposed model for robust outcomes, we used k-fold cross-validation. In this method, k-1 folds are used for training, while 1 fold is used for testing. For the PHEME dataset, we set k to 9. This led us to use one event of the PHEME dataset for testing, while the remaining events were used for training the

Table 1 Statistics of datsets post pre-processing

Dataset	PHEME	SemEval
Incidents	9	10
Tweets	105,354	5568
Threads	2402	325
Thread depth	2.8	3.5
True rumors	1067	145
False rumors	638	74
Unverified rumors	697	106

proposed framework. Similarly, SemEval has tweets covering 10 events in 325 conversation threads. Following the same methodology as the PHEME dataset, we set k to 10, but this time, we used 2 events for testing and the remaining 8 for training the model. Upon a detailed analysis of both datasets, we concluded that both datasets have an issue of class imbalance. Therefore, to have a fair analysis and a wholesome comparison, we chose the $Macro - F_1$ score and accuracy as evaluation metrics in our case.

Comparative analysis

In this section, we will explore the primary reasons for the outperformance of the proposed model. To conduct a comparative analysis with other state-of-the-art baselines, we have selected the following frameworks proposed in the literature:

- BranchLSTM [21]: It considers successive branches in a discussion thread and utilizes an LSTM-based framework for classifying the stance of rumors.
- TD-RvNN [4]: A recursive neural network framework driven by top-down propagation is used for rumor detection on social media.
- Hierarchical GCN-RNN [22]: The joint venture between graph convolutional and recurrent neural networks leverages the sequential and structural properties of conversational threads.
- PLAN [23]: A model based on a randomly initialized transformer is used to encode conversational threads for rumor detection.
- Hierarchical Transformer [24]: An extended BERT-based framework is proposed that learns the sub-thread interactions, followed by encoding their global interactions of all posts. The proposed model captures these interactions based on a Transformer layer.
- Bi-GCN [6]: A GCN-based model which formulates high-level representations on the bases of bottom-up and top-down views of conversation threads.
- ClaHi-GAT [25]: A GCN-based model formulates high-level representations based on both the bottom-up and top-down views of conversation threads.
- EBGCN [7]: Bi-GCN variant that adjust weights of unreliable relations through Bayesian method.

The proposed models used for detecting rumors can broadly examine conversation threads in two different aspects: structure-wise and branch-wise. Based on experimentation, it can be concluded that structure-wise exploitation of conversation threads gives better outcomes compared to branch-wise. We further validated this assumption by conducting a detailed comparative analysis of BranchLSTM with other models. BranchLSTM decomposes the conversation thread into branches of the tree and then encodes each branch to learn its feature representation. However, since LSTM is well-known for sequential data processing, it misses out on the abstract level representation of rumors that is embedded into the structural analysis of the thread. On the other hand, frameworks like Hierarchical Transformer, PLAN, and BiGCN evaluate the structural information of the conversation threads and perform better than BranchLSTM. Such models learn the structural representation of conversation threads which are critical for rumor detection. The reason for this criticality is such that the propagation of information in a social media platforms follow specific pattern.

Detailed analysis of Table 2 leads us to the conclusion that the performance of deep learning models is also affected by the perspective from which the conversation thread is analyzed. It is evident from Table 2 that EBGCN, Bi-GCN and ClaHi-GAT perform better among the frameworks that analyze the structural information of the conversation threads. These frameworks consider only a single perspective of the conversation threads, thus learning only the singleton view of the conversation and leaving out the reciprocal viewpoint. Despite the single perspective analysis, ClaHi-GAT outperforms its two competitors. The probable reason for this could be the attention heads (post-based, graph-based, and event-based) employed in ClaHi-GAT. The complex attention mechanism can extract meaningful information from conversation threads that can be helpful in detecting rumors in real-world scenarios.

RPf-GCN outperforms its baseline frameworks. The core reason for its better performance is that during training, it learns the reciprocal perspectives of conversation threads. This enables the proposed framework to learn indicators from multiple views of both the source and the respondents. By fusing these reciprocal perspectives, the proposed model can obtain a comprehensive view of rumors from a global standpoint, leading to a significant enhancement in the performance of our model. These observations indicate that by incorporating the reciprocal perspective structural data of conversation threads, our suggested model can adeptly identify rumors in real scenarios.

Table 2 Performance comparison

Framework	PHEME		SemEval	
	Macro-F1	**Acc**	**Macro-F1**	**Acc**
BranchLSTM [21]	0.491	0.500	0.259	0.314
TD-RvNN [4]	0.509	0.536	0.264	0.341
Hierarchical GCN-RNN [22]	0.540	0.536	0.317	0.356
PLAN [23]	0.581	0.571	0.361	0.438
Hierarchical Transformer [24]	0.592	0.607	0.372	0.441
Bi-GCN [6]	0.607	0.617	0.316	0.442
ClaHi-GAT [25]	0.539	0.536	0.369	0.556
EBGCN [7]	0.639	0.643	0.375	0.521
RPf-GCN (Proposed)	0.736 ±0.03	0.748 ±0.06	0.461 ±0.07	0.658 ±0.06

Ablation study

Modular analysis

In this subsection, we conducted an ablation study to examine the effectiveness of each component in RPf-GCN. We removed each component from the entire model and assessed its impact on the overall performance of the proposed framework. The term "Combined" denotes the complete model with all of its sub-modules. The ablated models include (1) "–Conv," which is RPf-GCN without the CNN. This approach is similar to the ones used in Bi-GCN and EBGCN, where the mean-pooling operation is applied to the $p \rightarrow q$ and $q \rightarrow p$ GCN to get their representations, followed by concatenation of both features for prediction. (2) "–$(q \rightarrow p)$," which is RPf-GCN that does not cater to the response-source perspective of the conversation thread. Similarly, (3) "–$(p \rightarrow q)$," is the variant of the proposed framework, which considers the source-response perspective of the thread. (4) "–Dir," where the conversation thread is modeled as an undirected tree structure encoded by a two-layer GCN added with a CNN submodule. (5) "GCN," which is the basic version of GCN, i.e., RPf-GCN without considering the reciprocal perspective.

Conclusions can be drawn from Table 3. First of all, it is evident that $RPf-GCN_{-(Dir)}$ performs better than both of its variants, which only consider the single perspective of conversation threads, i.e., $RPf-GCN_{-(p\rightarrow q)}$ and $RPf-GCN_{-(q\rightarrow p)}$. But the proposed model $RPf-GCN$ with all modules combined outperforms three of the $RPf-GCN_{-(Dir)}$, $RPf-GCN_{-(p\rightarrow q)}$ and $RPf-GCN_{-(q\rightarrow p)}$. This validates that taking into account both $p \rightarrow q$ and $q \rightarrow p$ reciprocal views lead to superior performance of the model.

Secondly, $RPf-GCN_{-(p\rightarrow q)}$ experiences a significant drop in performance compared to $RPf-GCN_{-(q\rightarrow p)}$, indicating that the source-to-respondent $(p \rightarrow q)$ propagation perspective is better at reflecting the characteristics of rumors than the respondent-to-source $(q \rightarrow p)$ dispersion view.

Thirdly, when the Conv component is removed, both RPf-GCN and $RPf-GCN_{-(Dir)}$ experience severe drops in Macro-F1 and Acc on both datasets. This demonstrates the effectiveness of the Convolutional module in feature representation for rumor detection. It not only fuses the reciprocal perspective information effectively but also captures enriched features for identifying rumors while considering only one perspective.

Table 3 Performance comparison with and without different sub-modules

Modules	**PHEME**		**SemEval**	
	Macro-F1	**Acc**	**Macro-F1**	**Acc**
Combined	0.736	0.748	0.461	0.658
–Conv	0.578	0.579	0.295	0.403
–$(q \rightarrow p)$	0.701	0.7	0.405	0.643
–$(p \rightarrow q)$	0.421	0.471	0.381	0.547
–Dir	0.525	0.536	0.405	0.556
GCN	0.446	0.493	0.371	0.547

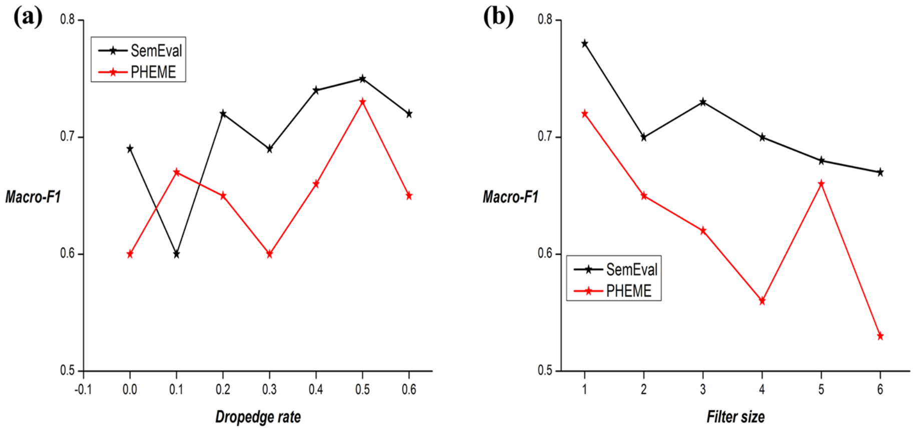


Fig. 3 Performance of RPf-GCN **a** Effect of Dropedge rate, **b** Effect of Filter size

Filter size analysis

We conducted an experiment to analyze the impact of varying the filter size in the Conv sub-module on rumor detection. Figure 3b displays the plot of macro-F1 score against various filter sizes, revealing that our suggested model achieved optimal performance on both datasets with a window size of 1. As the filter size increased, performance initially dropped, followed by marginal improvement with an increase in filter size. This aligns with our intuition that unlike the relationship between adjacent pixels in an image, there may not be a direct correlation between posts in chronological order. Thus, a larger window size led the model to learn more noise that hindered its performance. However, increasing the window size slightly enhanced the correlation between users, resulting in some improvement in the model's performance. Additionally, since there are few participating users and contents in the early stages of rumor propagation, a smaller window size was more effective for early stages of rumor detection. Consequently, the proposed framework holds good for early rumor detection.

Drop rates effect

In Fig. 3a, we tested the performance of RPf-GCN by varying the dropedge from 0 to 0.6. The performance showed a gradual increase, peaking at 0.5 before subsequently declining. Conversation threads often contain unreliable relationships, resulting in significant error accumulation that decreases model robustness [7]. Increasing the rate of dropping the edge leads to a decrease in the number of unreliable edges, improving model performance and robustness by enabling it to learn more compelling features. However, clipping too many edges ultimately leads to a decline in performance. Based on our experimentation, we determined that this reasonable rate produces the best results.

Conclusion

Social media is exploited by anti-state agents and state-sponsored groups for disinformation campaigns with political and strategic objectives. Detecting malicious content on social media is crucial. This paper introduces a novel deep learning framework based on a reciprocal perspective-driven graph convolutional neural network to effectively detect social media rumors. It treats rumor conversation threads as color images, integrating source and follower perspectives as channels and graph nodes as pixels. The model uses two concurrent GCNs to capture discriminating features from each perspective. A convolutional operation captures consistent and complementary information, resulting in a comprehensive conversation representation. Experimental results on real-world datasets show that our RPf-GCN significantly outperforms existing methods. The core reason for its superior performance is its ability to learn reciprocal perspectives, providing a comprehensive view of rumors and enhancing overall model performance.

Author contributions

ZK, JG and NI contributed to proposing the idea, conceptualization, methodology, experimentation, implementation, result compilation, writing-original draft preparation and drawing of figures etc. WP and MJ contributed to reviewing, editing and supervision.

Funding

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) Grant funded by the Korea Government (MSIT) (No. 2014-3-00077) and National Research Foundation of Korea(NRF) Grant funded by the Korea government (MSIT) (No. 2023R1A2C2006264).

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 19 June 2023 Accepted: 14 December 2023
Published online: 09 January 2024

References

1. Rubin VL. On deception and deception detection: content analysis of computer-mediated stated beliefs. Proc Am Soc Inf Sci Technol. 2010;47(1):1–10.
2. DiFonzo N, Bordia P. Rumor, gossip and urban legends. Diogenes. 2007;54(1):19–35.
3. Qazvinian V, Rosengren E, Radev D, Mei Q. Rumor has it: identifying misinformation in microblogs. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. 2011:pp. 1589–1599.
4. Ma J, Gao W, Wong KF. Rumor detection on twitter with tree-structured recursive neural networks 2018.
5. Li Q, Zhang Q, Si L. Rumor detection by exploiting user credibility information, attention and multi-task learning. In: Proceedings of the 57th annual meeting of the association for computational linguistics. 2019:pp. 1173–1179.
6. Bian T, Xiao X, Xu T, Zhao P, Huang W, Rong Y, Huang J. Rumor detection on social media with bi-directional graph convolutional networks. In: Proceedings of the AAAI conference on artificial intelligence. 2020:pp. 549–556.
7. Wei L, Hu D, Zhou W, Yue Z, Hu S. Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection. 2021. arXiv preprint arXiv:2107.11934.
8. Ozbay FA, Alatas B. Fake news detection within online social media using supervised artificial intelligence algorithms. Phys A Stat Mech Appl. 2020;540: 123174.
9. Kaliyar RK, Goswami A, Narang P, Sinha S. FNDNet-a deep convolutional neural network for fake news detection. Cogn Syst Res. 2020;61:32–44.
10. Faustini PHA, Covões TF. Fake news detection in multiple platforms and languages. Expert Syst Appl. 2020;158: 113503.
11. Wang Y, Wang L, Yang Y, Lian T. SemSeq4FD: integrating global semantic relationship and local sequential order to enhance text representation for fake news detection. Expert Syst Appl. 2021;166: 114090.
12. Liu Y, Wu Y. FNED: a deep network for fake news early detection on social media. ACM Trans Inf Syst. 2020;38(3):1–33.
13. Ma J, Gao W, Wei Z, Lu Y, Wong KF. Detect rumors using time series of social context information on microblogging websites. In: Proceedings of the 24th ACM international on conference on information and knowledge management. 2015;pp. 1751–1754.
14. Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong KF, Cha M. Detecting rumors from microblogs with recurrent neural networks 2016.

15. Liu Y, Wu YF. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Proceedings of the AAAI conference on artificial intelligence 2018.
16. Lu YJ, Li CT. Gcan: Graph-aware co-attention networks for explainable fake news detection on social media. arXiv preprint arXiv:2004.11648 2020.
17. Benevenuto F, Magno G, Rodrigues T, Almeida V. Detecting spammers on twitter. In Collaboration, electronic messaging, anti-abuse and spam conference (CEAS). 2010;6:12.
18. Thomas K, McCoy D, Grier C, Kolcz A, Paxson V. Trafficking fraudulent accounts: the role of the underground market in twitter spam and abuse. in Presented as part of the 22nd fUSENIXg Security Symposium (fUSENIXg Security 13), 2013: pp. 195–210.
19. Jiang M, Cui P, Beutel A, Faloutsos C, Yang S. Detecting suspicious following behavior in multimillion-node social networks. in Proceedings of the 23rd International Conference on World Wide Web, 2014: pp. 305–306.
20. Rong Y, Huang W, Xu T, Huang J. The truly deep graph convolutional networks for node classification. CoRR abs/1907.10903 2019. arXiv:1907.10903.
21. Kochkina E, Liakata M, Augenstein I. Turing at semeval-2017 task 8: Sequential approach to rumour stance classification with branch-lstm. arXiv preprint arXiv:1704.07221 2017.
22. Wei P, Xu N, Mao W. Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity. arXiv preprint arXiv:1909.08211 2019.
23. Khoo LMS, Chieu HL, Qian Z, Jiang J. Interpretable rumor detection in micro-blogs by attending to user interactions. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2020:pp. 8783–8790.
24. Yu J, Jiang J, Khoo LMS, Chieu HL, Xia R. Coupled hierarchical transformer for stance-aware rumor verification in social media conversations 2020.
25. Lin H, Ma J, Cheng M, Yang Z, Chen L, Chen G. Rumor detection on twitter with claim-guided hierarchical graph attention networks. arXiv preprint arXiv:2110.04522 2021.
26. Zhang Y, Yu X, Cui Z, Wu S, Wen Z, Wang L. Every document owns its structure: Inductive text classification via graph neural networks. arXiv preprint arXiv:2004.13826 2020.
27. Wei Y, Wang X, He X, Nie L, Rui Y, Chua TS. Hierarchical user intent graph network for multimedia recommendation. IEEE Transactions on Multimedia 2021.
28. Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. arXiv preprint arXiv:1609.02907, 2018.
29. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks stat. 2017;1050:20.
30. Xie Y, Zhang Y, Gong M, Tang Z, Han C. Mgat: multi-view graph attention networks. Neural Netw. 2020;132:180–9.
31. Castillo C, Mendoza M, Poblete B. Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web. 2011:pp. 675–684.
32. Feng S, Banerjee R, Choi Y. Syntactic stylometry for deception detection. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2012:pp. 171–175.
33. Chen Y, Conroy NJ, Rubin VL. Misleading online content: recognizing click bait as " false news". In: Proceedings of the 2015 ACM on workshop on multimodal deception detection. 2015:pp. 15–19.
34. Kwon S, Cha M, Jung K, Chen W, Wang Y. Prominent features of rumor propagation in online social media. In: 2013 IEEE 13th international conference on data mining. pp. 1103–1108. IEEE 2013.
35. Sampson J, Morstatter F, Wu L, Liu H. Leveraging the implicit structure within social media for emergent rumor detection. In: Proceedings of the 25th ACM international on conference on information and knowledge management. 2016: pp.2377–2382.
36. Yang F, Liu Y, Yu X, Yang M. Automatic detection of rumor on sina weibo. In: Proceedings of the ACM SIGKDD workshop on mining data semantics. 2012; pp. 1–7.
37. Peng Z, Zhen H, Yong D, Yeqing Y. Rumor detection on social media through mining the social circles with high homogeneity. Inf Sci. 2023;642: 119083.
38. Lin H, Yi P, Ma J, Jiang H, Luo Z, Shi S, Liu R. Zero-shot rumor detection with propagation structure via prompt learning. Proc AAAI Conf Artif Intell. 2023;37(4):5213–21. https://doi.org/10.1609/aaai.v37i4.25651.
39. Li S, Li W, Luvembe AM, Tong W. Graph Contrastive Learning With Feature Augmentation for Rumor Detection. in IEEE Transactions on Computational Social Systems, https://doi.org/10.1109/TCSS.2023.3269303.
40. Sun M, Zhang X, Ma J, Xie S, Liu Y, Philip SY. Inconsistent Matters: A Knowledge-guided Dual-consistency Network for Multi-modal Rumor Detection. IEEE Transactions on Knowledge and Data Engineering. 2023.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.