

Article

Rumor Detection on Social Media via Fused Semantic Information and a Propagation Heterogeneous Graph

Zunwang Ke ¹, Zhe Li ¹, Chenzhi Zhou ², Jiabao Sheng ¹ and Wushour Silamu ^{3,*}
and Qinglang Guo ⁴

¹ College of Software, Xinjiang Laboratory of Multi-Language Information Technology, Xinjiang Multilingual Information Technology Research Center, Xinjiang University, Urumqi 830046, China; kzwang@xju.edu.cn (Z.K.); lizhe@stu.xju.edu.cn (Z.L.); jiabao@stu.xju.edu.cn (J.S.)

² Shukutoku Japanese Language School, Tokyo 197-0804, Japan; zcz1997@gmail.com

³ Xinjiang Multilingual Information Technology Research Center, Xinjiang Laboratory of Multi-Language Information Technology, College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China

⁴ China Academy of Electronics and Information Technology, National Engineering Laboratory for Risk Perception and Prevention (NEL-RPP), Beijing 100041, China; gq1993@mail.usc.edu.cn

* Correspondence: wushour@xju.edu.cn

Received: 13 October 2020; Accepted: 30 October 2020; Published: 31 October 2020



Abstract: Social media had a revolutionary impact because it provides an ideal platform for share information; however, it also leads to the publication and spreading of rumors. Existing rumor detection methods have relied on finding cues from only user-generated content, user profiles, or the structures of wide propagation. However, the previous works have ignored the organic combination of wide dispersion structures in rumor detection and text semantics. To this end, we propose KZWANG, a framework for rumor detection that provides sufficient domain knowledge to classify rumors accurately, and semantic information and a propagation heterogeneous graph are symmetry fused together. We utilize an attention mechanism to learn a semantic representation of text and introduce a GCN to capture the global and local relationships among all the source microblogs, reposts, and users. An organic combination of text semantics and propagating heterogeneous graphs is then used to train a rumor detection classifier. Experiments on Sina Weibo, Twitter15, and Twitter16 rumor detection datasets demonstrate the proposed model's superiority over baseline methods. We also conduct an ablation study to understand the relative contributions of the various aspects of the method we proposed.

Keywords: fake news detection; attention; GCN; heterogeneous graph; social networks

1. Introduction

With the rapid development of large-scale social network platforms such as Sina Weibo, Jinri Toutiao, and Tik Tok, rumor identification on social media has been a challenging topic. Rumors can spread and affect people's opinions due to the convenience of social media; however, rumors can cause significant harm to society and can result in huge economic losses. Therefore, to address the potential of rumors causing panic and threats, it is of high practical significance to propose a method that can efficiently identify rumors in social media content.

Previous research on automatic rumor detection has concentrated largely on extracting effective features from many different types information sources, including text content [1–3], user profiles [1,4], and propagation patterns [5–7]. However, the complexity and scale of social media data pose a sea of technical challenges. First, social media language is casual and informal, usually dynamic or

ungrammatical; thus, traditional NLP techniques cannot be applied directly. Second, when using handcrafted features, there are many instances where individual or multiple handcrafted features are either unavailable, inadequate, or manipulated. Inspired by recent achievements in deep learning, the more recent studies have applied different neural networks to rumor detection tasks, including rumor detection itself [8], the identification of candidate rumors, and rumor verification [9,10]; the goal of the latter is to assess the veracity of a rumor. However, among different users and microblogs, these models ignore the information of global structural which has been verified to be useful in providing helpful clues for node classification [7].

Rumor structures spread also indicate that rumors have some particular spreading behaviors. Therefore, some studies have tried to include information regarding rumor structures dispersion by invoking CNN-based methods [11,12]. CNN-based methods can extract correlated characteristics among local neighbors but cannot deal with structural relationships which exist in trees or graphs [13]. Therefore, these approaches ignore rumor dispersion structural features. In fact, a CNN is not designed to study representations of high-level from structured data, but the graph convolutional network (GCN) can do [14]. Social networks are typically structured as heterogeneous graphs containing entities such as users, posts, geographic locations, hashtags, and these denote relationships such as followers, friendships, retweets, and spatial neighborhoods. Therefore, these heterogeneous networks provide new and different perspectives regarding the relationships among microblogs; thus, they contain rich information that can improve the rumor detection performance. However, most of the previous studies on rumor detection have treated each source microblog as independent, so that one does not affect another. Thus, they have not fully exploited the correlations between different node types.

Therefore, in this study, we introduce GCN to capture the structural information of spreading and diffusion in all the source microblogs, reposts, and users. To clarify our motivation, Figure 1 shows a global heterogeneous graph containing three source microblogs with corresponding user responses. In this example, two users ($User_1$ and $User_2$) have no friendship relations, they do not follow each other, but they both repost the same Weibo content $weibo_1$. In addition, the three Weibo posts are not related in content, but $weibo_2$ and $weibo_3$ share neighbors are similar, which shows they probably have the same tags. We build a global heterogeneous graph over these observations to capture the local and global relationships between all the Weibo sources, reposts, and users. To achieve rumor detection, we first learn word embeddings from the text contents of microblogs posted by users. Then, we use a GCN to learn the representation of repost propagation and learn the representation of user interactions via graph-aware. Then, we construct a graph to model the potential interactions between users.

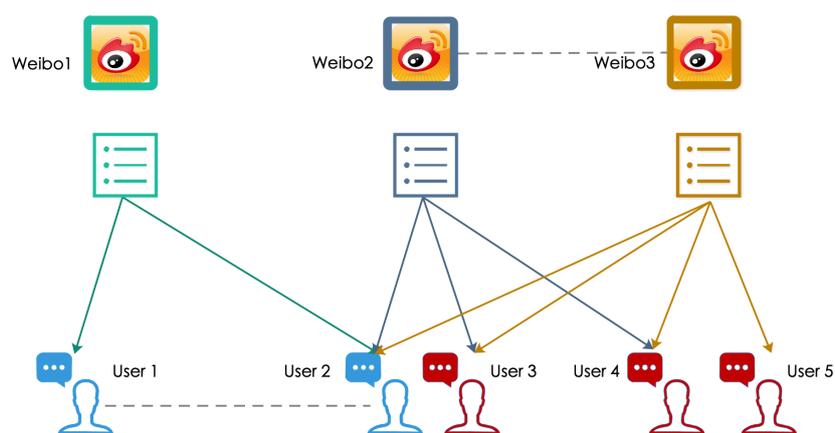


Figure 1. An example of rumor propagation in a heterogeneous graph.

In the method, we proposed symmetry fusing the semantic information and a propagation heterogeneous graph. The method we proposed also achieves much better effectiveness in early rumor detection, which is quite powerful to identify rumors and avoid rumor spreading.

The main contributions of this work are as follows:

- We utilize a deep integration of rumor propagation along relationship chains and text semantic information via the heterogeneous network to detect rumors.
- We apply multihead attention to fuse the local semantic information to generate a better-integrated representation for each microblog.
- We concatenate the source microblog features with other microblogs at each graph convolutional layer to comprehensively use the root feature information and obtain excellent rumor detection performance.

2. Related Works

Automatic rumor detection aims to identify whether a microblog text on a social networking platform is a rumor via its related information, including comments, text content, communication mode, propagation patterns, etc. Ref. [15] has provided a survey of research into social media rumors. The related works that have proposed the most recent techniques to date can be grouped into three main categories: (1) deep learning methods, (2) graph neural networks, and (3) propagation tree-related methods.

2.1. Deep Learning Methods

Recent deep learning advances have successfully attained state-of-the-art performances on natural language processing tasks. Researchers have applied deep learning models to automatically learn efficient characteristics for detecting rumors. Ref. [16] presented an innovative method that learns successive representations of microblog events to identify rumors; the proposed model was based on a recurrent neural network (RNN), which learns hidden representations that capture the variations in the contextual information of relevant posts that occur over time. Ref. [9] proposed a novel model with a shared layer and two task-specific layers. They incorporated the information of user credibility into the rumor identification layer. In this study, we also introduce an attention mechanism to the rumor identification process. Ref. [17] utilized a specific task character based on the bidirectional language model and stacked LSTM to represent the text information and social temporal information of source tweets inputted to model the disseminate patterns of rumors during the development early stages. They also introduced multilayered attention models to jointly learn context embeddings via many context inputs. However, the above methods not only ignore the spread mode of microblog, but also model the diffusion path as a well-aligned structure, and fail to make the best of the diffusion information of microblog. Moreover, these methods focus little on the discovery of earlier rumors.

2.2. Graph Neural Network

Traditional deep learning methods only consider the patterns of deep propagation; they ignore the structures of wide dispersion during rumor detection. To address this challenge, a bi-directional graph convolutional network is proposed to explore these two features through top-down and bottom-up rumor spreading [18]. Their method is to use GCN and top-down rumor spreading maps to learn the pattern of rumor spreading, and use GCN with rumor spreading maps in opposite directions to capture the structure of rumor spreading. Ref. [19] proposed the graph-aware co-attention network (GCAN), which uses RNN and convolutional to learn a representation of repost diffusion via user features. Firstly, a graph is constructed to simulate the potential interaction between users, and then GCN is used to learn the graphical perception representation of user interaction. A new global-local attention network (GLAN) for detecting rumor is proposed [20], which encodes local semantics and global structure information jointly. In this study, we first fuse the semantic information of related tweets with the attention mechanism to produce a better-integrated representation for every content of the tweet. Then, we model the relationships in all source microblog, forwards, and users as heterogeneous graphs to obtain rich information of structural for identify rumors.

2.3. Propagation Tree Related Methods

Different from previous works that paid attention to microblog text content, the diffusion of tree-related methods pays attention to the differences in the features of real and fake information transference. Ref. [21] proposed to learn discriminative features from microblog posts via following their non-sequential diffusion structure to generate more strong representations for detecting rumors. They use a propagation tree to represent the spread of micro-blog posts, which provides useful clues as to how the claims in the original posts spread and develop over time. Ref. [8] utilized two recursive neural models based on top-down and bottom-up structured of tree neural networks to represent rumor learning and classification and tried to learn discriminative features from tweet content by following their nonsequential propagation structure and generating more helpful representations for detecting different types of rumors. Ref. [6,7,22] applied propagation tree methods to rumor detection, providing useful clues on how a microblog diffuses and develops over time.

Message proliferation ahead of online networking is essentially spread in the structure in the heterogeneous chart. In the heterogeneous graph, users convey or alternately forward messages to propagate quickly and widely. These techniques are dependent upon the proliferation trees used to investigate the contrasts in the structure of data transmission; however, it has not been considered for the connections around the distinctive proliferation trees.

3. Problem Statement

Let $\mathcal{M} = \{m_1, m_2, m_3, \dots, m_i\}$ be a set of source microblogs and $\mathcal{U} = \{u_1, u_2, u_3, \dots, u_i\}$ be a set of users. Each source microblog m_i consists of n repost microblogs $\{r_1, r_2, \dots, r_n\}$. For every microblog, we use the notation t_i to denote the post time. W repost microblogs as neighbor nodes of the twitter or microblog in the heterogeneous graph, which are formulated as $\mathcal{N}(m_i) = [r_1, r_2, \dots, r_n]$.

We aim to train a function $P(c = 1 | m_i, \mathcal{N}(m_i); \mathcal{U}; \theta)$, it can predict a twitter or Sina Weibo is a rumor or not. Here, c is the label of class, and θ represents all the model parameters.

4. Methodology

The rumor detection model proposed in this paper consists of three main parts: microblog representation, representations of propagation and dispersion, and rumor detection. The module of microblog representation depicts microblogs mapping from word embedding to the space of semantic. Propagation and decentralized representations use a GCN to describe user propagation representations. The module of rumor detection train a function of classification, the function can predict the tags of the Twitter or Sina Weibo. As shown in Figure 2, it is the framework of KZWANG whose model we proposed. In the following, we introduce each major component in detail.

4.1. Microblog Representation

We utilize the multi-head attention [23] to represent microblog via learn context of Twitter or Sina Weibo. The module of multi-head attention uses three sentences of input: a sentence of query, a sentence of key, and a sentence of value, namely, $Q \in \mathbb{R}^{n_q \times d}$, $K \in \mathbb{R}^{n_k \times d}$, and $V \in \mathbb{R}^{n_v \times d}$, respectively, where n_q , n_k , and n_v denote the amount of words in every sentence, and d is the embedding dimension. Attention module is the most significant module in the coding unit, and it can be defined as follows:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

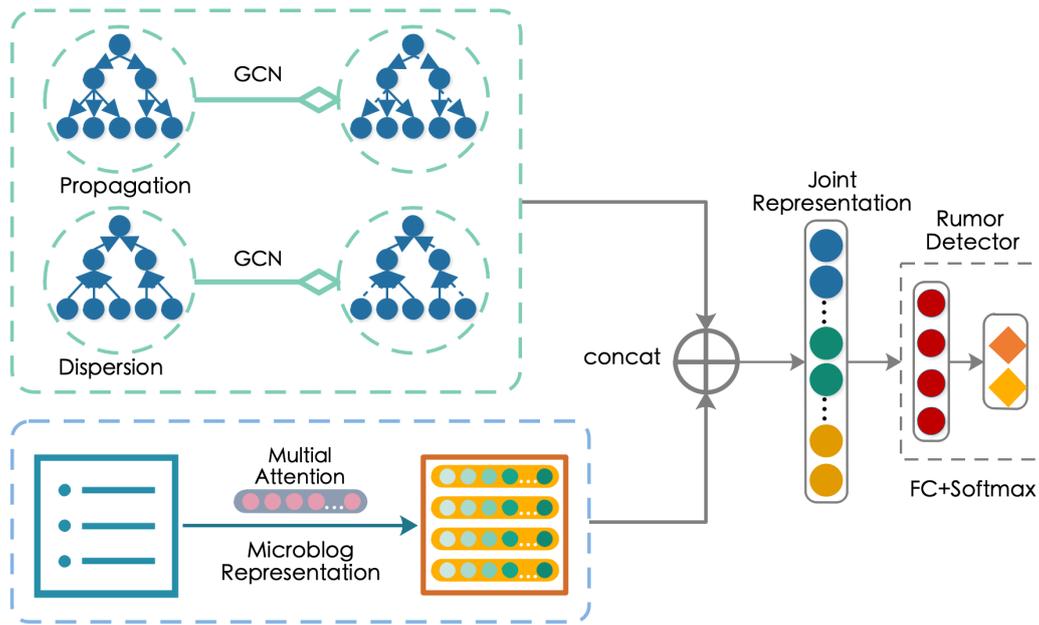


Figure 2. The architecture of the proposed rumor detection model.

By this means, the attention does catch dependencies among query sentences and key sentences and further use the relation information to assemble the components in the sentence of query and convert value sentences into component representations. To extend the model's ability to focus on different positions and improve the representation learning ability of the attention unit subspace, the transformer applies a "multi-head" mode that can be expressed as follows:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (2)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^K) \quad (3)$$

We obtained microblog representations from word embeddings in the same way.

4.2. Propagation and Dispersion Representation

4.2.1. Construct Propagation and Dispersion Graphs

We aim to create a graph to model the potential interactions among users who repost the source story. The idea is that the correlations between users with specific features can reveal the feasibility that the Weibo post is a rumor. Specifically, the propagation graph contains the source Weibo node (called a "Weibo" hereafter), the publisher node of the source Weibo, and the repost node. The graph weight is calculated based on the time difference between the repost Weibo and the main Weibo, and the calculation method is $\frac{1}{1+i}$. An edge represents a publishing relationship and a repost relationship. Then, all the nodes are uniformly numbered so that the adjacency matrix contains all the relationships.

4.2.2. Propagation and Dispersion Encoding

GCN is a multi-layer neural network, which processes graph data and generates embedded vectors of nodes according to their neighborhoods. GCN is able to catch information from a node direct and indirect neighbors via stacked hierarchical convolution. Many message propagation functions types, M , exist for GCN [13,24], where the message propagation function defined in ChebNet's [14] first-order approximation is as follows:

$$H_k = M(A, H_{k-1}; W_{k-1}) = \sigma(\hat{A}H_{k-1}W_{k-1}) \quad (4)$$

In the above formula, the $H_k \in \mathbb{R}_n \times vk$ is the hidden feature matrix, and M is the message diffusion function, and it relies on the adjacency matrix A , the hidden feature matrix H_{k-1} , and the trainable parameters W_{k-1} . \hat{A} is the normalized adjacency matrix, and $\sigma(\cdot)$ is an activation function.

4.2.3. Root Feature Enhancement

As is known to all, the original information of a rumor event usually contains a lot of information, so it can have a broad effect. It is important to make better use of the source thread info and learn more precise node representations from the relationship between source and nodes. Therefore, besides the hidden features from TD-GCN and BU-GCN, we introduce a root feature enhancement operation to enhance rumor detection performance, as shown in Figure 3,

$$\hat{H}_k^{TD} = \text{concat}(H_k^{TD}, (H_{k-1}^T D)^{root}) \tag{5}$$

with $H_0^{TD} = X$. Therefore, we express TD-GCN using root feature enhancement via substituting H_1 in Equation (5) with $\hat{H}_1^{TD} = \text{concat}(H_1^{TD}, X^{root})$, and then obtain \hat{H}_2^{TD} as follows:

$$H_2^{TD} = \sigma(\hat{A}^{TD} \hat{H}_1^{TD} W_1^{TD}) \tag{6}$$

$$\hat{H}_2^{TD} = \text{concat}(H_2^{TD}, (H_1^{TD})^{root}) \tag{7}$$

Similarly, as shown in Figure 4, BU-GCN hidden feature measures based on root feature enhancement, \hat{H}_1^{BU} and \hat{H}_2^{BU} , are obtained in the same manner as Equations (6) and (7).

4.3. Rumor Classification

After the above procedures, we have obtained a text representation t_j and a propagation representation p_j . These two methods of representation are very important for rumor detection. Therefore, these two methods are related to each other to form the final characteristics for classification. Then, the final representation is projected into the target probabilistic space using the fully connected layer:

$$p(c|m, N(m), U; \theta) = \text{softmax}(W[t; p] + b) \tag{8}$$

where $W \in R^{2d \times |c|}$ is the weight parameter and $b \in R$ is a bias term.

Finally, for the optimization objective function, we adopt cross-entropy loss to identify rumor:

$$J(c^i|D, u^i; \theta) = - \sum y_i \log p_i(c|m_i, N(m_i), U; \theta) \tag{9}$$

where y_i is the gold-standard probability of the class of rumor and θ represents all the model parameters.

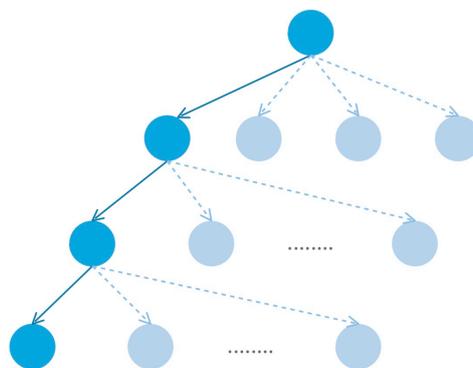


Figure 3. TD-GCN: the deep propagation along a relationship chain from top to bottom.

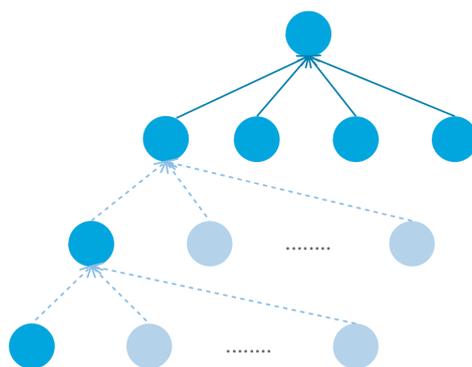


Figure 4. BU-GCN: the aggregation of the wide dispersion within a community to an upper node.

5. Experiments

Firstly, we assess KZWANG model empirical performance and compare it with several baseline models. Then, we do ablation studies to explore the influence of the essential elements of the KZWANG model. Finally, we also test the ability of both KZWANG and the compared methods to perform early rumor detection. Replicating the experiment code is available at <https://github.com/shanmon110/rumordetection>.

5.1. Datasets

The experiments on the KZWANG model has been done on three social media datasets: Twitter15, Twitter16 [7], and Sina Weibo [16], from the most popular social media sites in the United States and China, respectively. The Sina Weibo dataset is annotated with binary labels, either “false rumor” or “non-rumor”. The Twitter16 datasets are each annotated with four tags, i.e., “false rumor” (FR), “non-rumor” (NR), “unverified rumor” (UR), and “true rumor” (TR). The label “true rumor” denotes a Twitter that tells people that a certain Twitter is a rumor. About every dataset, we can build heterogeneous graphs via the source tweet, the response tweet, and the relevant users. Table 1 shows summary statistics for the three datasets.

Table 1. Topic classification accuracy results.

Statistic	Weibo	Twitter15	Twitter16
# Source tweets	4664	1490	818
# Non-rumors	2351	374	205
# False rumors	2313	370	205
# Unverified rumors	0	374	203
# True rumors	0	372	205
# Users	2,746,818	276,663	173,487
# Posts	3,805,656	331,612	204,820

5.2. Baselines

We compare our method with the following baseline rumor detection models:

- **DTC** [1] A model based on decision tree that uses a combination of news features.
- **SVM-RBF** [4] An SVM model with an RBF kernel that introduces a combination of news features.
- **SVM-TS** [25] An SVM model that introduces a time sequence to model the variations of news features.
- **DTR** [26] Detection and ranking method of query phrase rumor based on decision tree.
- **GRU** [16] Based on the RNN model, the temporal language pattern is studied from user comments.
- **RFC** [27] A random forest classifier via utilizing linguistic, user, and structural features.

- **PTK** [7] An SVM classifier with diffusion tree kernel detects rumors by studying the time structure mode of propagation tree.
- **RvNN** [28] A top-down and bottom-up model based on tree structure via recursive neural networks for fake news identified on Twitter.
- **PPC** [29] A novel model for rumor detection by classifying propagation paths by a combination of recurrent and convolutional networks.
- **GLAN** [20] A novel rumor detection model with global-local attention network (GLAN) that jointly encodes global structural information and the local semantic.

5.3. Setup

We randomly selected 10% of the instances as a validation dataset and split the rest among the training datasets and testing datasets at a ratio of 3:1 for all three datasets. In the model, all the word embeddings are initialized with 300-dimensional word vectors that trained on a corpora of a specific domain, and it is trained by the skip-gram algorithm. Uniformly distributed initialization is performed for words that do not exist in the pre-trained word vector set. During training, we keep word vectors trainable and for each task they can be fine-tuned. In the Twitter15/16 datasets, the words are segmented with spaces, while the words in the Weibo dataset are split by the Jieba library. We removed words that had more than two occurrences since those words might be the stop words. Our model was implemented using PyTorch, and introduced Adam algorithm [30] to update parameters, with the following parameters: β_1 and β_2 are 0.9 and 0.999, we initialized learning rate is 1×10^{-3} and gradually decreases during the training process. We selected the best parameters on the base of model performance on the training dataset and evaluated the best parameters on the test datasets. The batch size of the training set was set to 64.

5.4. Evaluation Metrics

To evaluate the text classification performance of the model in this classification experiment, we introduced the precise rate (P), the recall rate (R), and the F1-metrics as the evaluation criteria. The calculation formulas for these indicators are as follows:

$$\left\{ \begin{array}{l} P = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \\ F1 = \frac{2 \times P \times R}{P + R} \end{array} \right. \quad (10)$$

In the above formula, TP represents the total number of sentences predicted as positive classes. FN means the total number of positive class sentences are predicted to be negative class. FP represents the total number of sentences whose negative class is forecasted to be positive label. TN means the total number of sentences predicted to be negative.

5.5. Results and Analysis

Tables 2–4 show all compared models performance. The baseline models experimental results are cited from previous studies [20] directly due to compare fair.

From Tables 2–4, we can see that the proposed KZWANG performs better than all other baselines on the Twitter and Sina Weibo datasets. Specifically, our model KZWANG attains an accuracy of 95.0% on the Sina Weibo dataset and accuracies of 91.1% and 90.7% on the two Twitter datasets. These results indicate the adaptability of KZWANG on different types of datasets. In addition, these excellent results show that using a GCN heterogeneous graph models can effectively learn node representations using both semantic and structural information.

The performances of the models based on handcrafted features (DTR, DTC, RFC, SVM-RBF, and SVM-TS) gained evidently poor performance, showed they are unsuccessful in generalizing across the datasets due to their inability to capture sufficient helpful characteristics. In the above baselines, SVM-TS and RFC obtained a relatively better performance since they utilize extra structural characteristic or temporal features, Still, they clearly perform worse than models that do not rely on feature engineering.

Compared the two diffusion models based on tree structure, PTK depends on structural characteristic extracted from propagation trees and linguistic. The RvNN-based model is inherently structure based on tree, and it capitalizes on the representation learning following propagation structure—so, it outperforms PTK. However, while modeling the propagation process, these methods based on tree miss too much information since the information is via a graph structure to spread, not a tree structure.

Among the deep learning models, GRU and PPC perform better than the traditional classifiers that use handcrafted features. This result shows neural network models that can learn representative deep latent characteristics automatically. Meanwhile, we can also notice that PPC is more better effective than GRU, due to the following reasons: (1) GRU depends on temporal-linguistic mode, while PPC depends on the user fixed features of forward sequences. (2) the PPC model unions CNN and RNN to catch the difference of user features.

Table 2. Experimental results on the Twitter15 dataset. The best and second-best results in each metric are bold and underlined, respectively.

Twitter15					
Method	Acc	NR F1	FR F1	TR F1	UR F1
DTR	0.409	0.501	0.311	0.364	0.473
DTC	0.454	0.733	0.355	0.317	0.415
RFC	0.565	0.810	0.422	0.401	0.543
SVM-RBF	0.318	0.455	0.037	0.218	0.225
SVM-TS	0.544	0.796	0.472	0.404	0.483
PTK	0.750	0.804	0.698	0.765	0.733
GRU	0.646	0.792	0.574	0.608	0.592
RvNN	0.723	0.682	0.758	0.821	0.654
PPC	0.842	0.811	0.875	0.818	0.790
GLAN	<u>0.905</u>	<u>0.924</u>	<u>0.917</u>	0.852	<u>0.927</u>
KZWANG	0.911	0.928	0.920	<u>0.850</u>	0.931

Table 3. Experimental results on the Twitter16 dataset. The best and second-best results in each metric are bold and underlined, respectively.

Twitter16					
Method	Acc	NR F1	FR F1	TR F1	UR F1
DTR	0.414	0.394	0.273	0.630	0.344
DTC	0.465	0.643	0.393	0.419	0.403
RFC	0.585	0.752	0.415	0.547	0.563
SVM-RBF	0.321	0.423	0.085	0.419	0.037
SVM-TS	0.574	0.755	0.420	0.571	0.526
PTK	0.732	0.740	0.709	0.836	0.686
GRU	0.633	0.772	0.489	0.686	0.593
RvNN	0.737	0.662	0.743	0.835	0.708
PPC	0.863	0.820	0.898	0.843	0.837
GLAN	<u>0.902</u>	<u>0.921</u>	0.869	<u>0.847</u>	0.968
KZWANG	0.907	0.926	<u>0.873</u>	0.850	0.971

Table 4. Experimental results on the Weibo dataset. The best and second-best results in each metric are bold and underlined, respectively.

Method	Class	Acc.	Prec.	Recall	F1
DTR	FR	0.732	0.738	0.715	0.726
	NR		0.726	0.749	0.737
DTC	FR	0.831	0.847	0.815	0.831
	NR		0.815	0.847	0.830
RFC	FR	0.849	0.786	0.959	0.864
	NR		0.947	0.739	0.830
SVM-RBF	FR	0.818	0.822	0.812	0.817
	NR		0.815	0.824	0.819
SVM-TS	FR	0.857	0.839	0.885	0.861
	NR		0.878	0.830	0.857
GRU	FR	0.910	0.876	0.956	0.914
	NR		<u>0.952</u>	0.864	0.906
PPC	FR	0.921	0.896	0.962	0.923
	NR		0.949	0.889	0.918
GLAN	FR	<u>0.946</u>	<u>0.943</u>	0.948	<u>0.945</u>
	NR		0.949	<u>0.943</u>	<u>0.946</u>
KZWANG	FR	0.950	0.945	<u>0.954</u>	0.949
	NR		0.954	0.945	0.950

In conclusion, KZWANG performs better than methods based on neural network and feature-based models, and its accuracy and precision on the rumor recognition task constitute improvements over previous models. Specifically, on the Weibo dataset, KZWANG improves the accuracy of the diffusion path classification model (the best baseline method) from 94.6% to 95.0%, and boosts the accuracy on the two Twitter datasets from 90.5% to 91.1% and from 90.2% to 90.7%, respectively. These results show the text information of semantic and the structure of propagation information are important to describe rumors and non-rumors difference.

5.6. Ablation Study

To determine the relative significance of each module in KZWANG, we conducted a series of ablation studies involving different modules of the model. The experimental results are shown in Figure 5. The ablation studies are conducted as follows:

- GCN: This experiment explored the efficacy using an LSTM or a GCN to encode the propagation graph for rumor classification.
- Attention: This experiment explored the efficacy of using an LSTM or an attention mechanism to extract text features from source tweets for rumor classification.
- Only Text: This experiment removed the propagation and dispersion encoding modules and used only text information for rumor classification.

Figure 5 shows the experimental results, we can observe the following. We first examined the influence of the GCN encoding module. We can see that replacing the GCN with an LSTM significantly affects the performance on all the datasets. The GCN captures the semantic relations between the source microblog and the corresponding repost microblogs; thus, these results show that it is critical to explicitly encode the propagation and dispersion relation.

Intuitively speaking, the attention mechanism causes the distance relevant semantic information to be close to each other, which leads to high cohesion and low coupling in the rumor and non-rumor groups, which improves the performance.

Finally, we evaluate the influence of the propagation heterogeneous graph encoding module. As shown in Figure 5, using only text results in significant performance declines on all the datasets.

In general, the performances obtained improve significantly after combining the text and propagation information, which shows that combining the information from these two aspects provides complementary effects from both local and global perspectives.

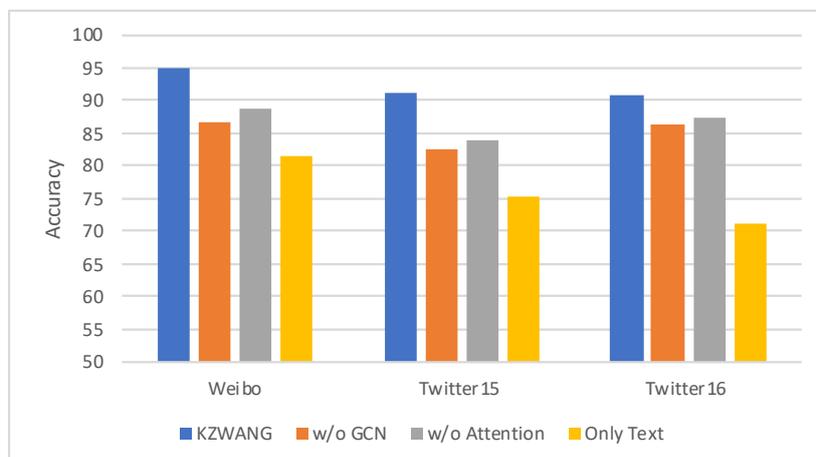


Figure 5. The ablation study results on the Weibo, Twitter15, and Twitter16 datasets.

5.7. Early Detection

Early detection aims to detect rumors at their early propagation stages—as early as possible—allowing interventions to be made quickly [26].

To construct an early detection task, we instituted deadlines for detection and used only published posts before those deadlines to evaluate the accuracy of the KZWANG method and the benchmark models. By changing the time delay for forwarding, several competitive models accuracy is shown in Figures 6–8. In the first few hours, our KZWANG model using only data within 4 h of the source microblog already outperforms the classification models based on tree structure (the best baselines) using the full data, which clearly indicates the superior early detection ability of our model. In particular, KZWANG achieves an accuracy of 92% on Weibo, 90% on Twitter15, and 85% on Twitter16 within the initial 4-hour period, which is dramatically faster than other benchmark methods.

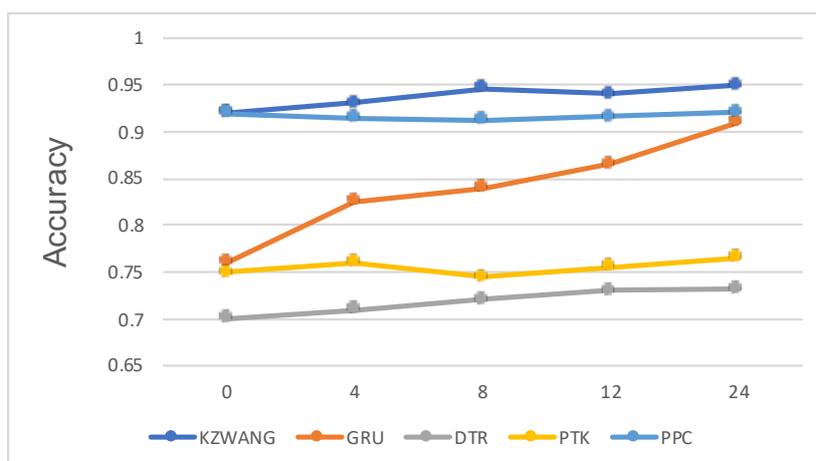


Figure 6. Weibo Detection Deadline (hours).

As we varied the time delays change 4 to be 12 h, KZWANG experienced a light drop, still it is superior to the state-of-the-art (SOAT) methods. The reason is as microblog propagate,

more information about semantic and structural is available. However, the noise will also increase. The results indicate that KZWANG is apathetic to data and it has fine robustness.

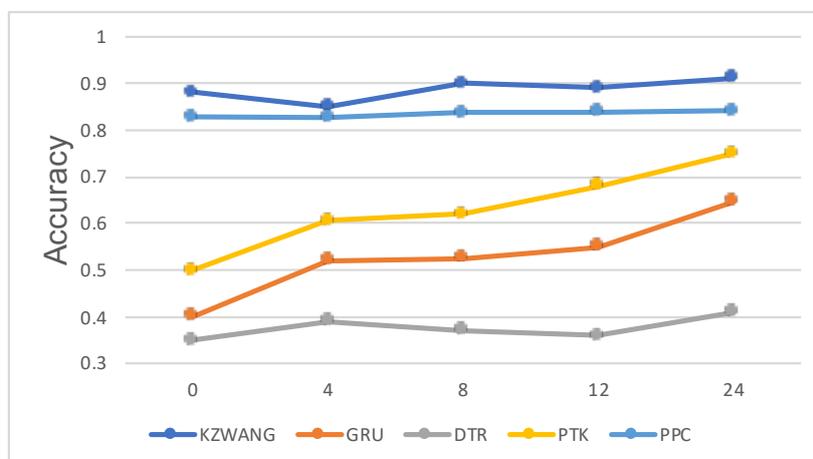


Figure 7. Twitter15 Detection Deadline (hours).

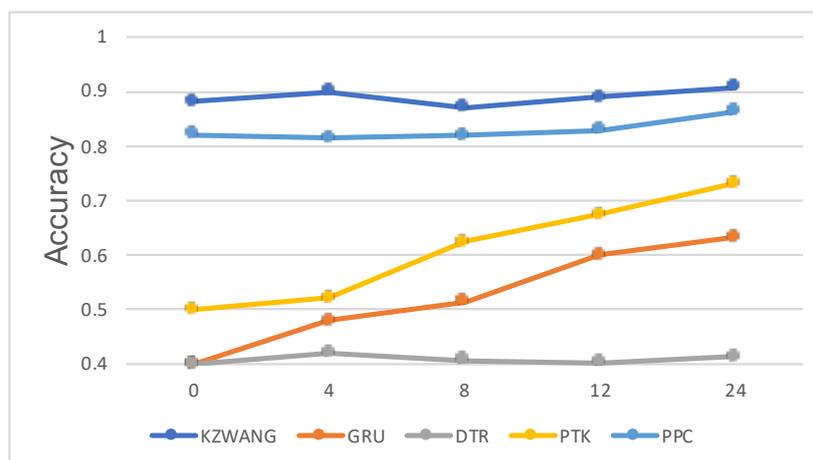


Figure 8. Twitter16 Detection Deadline (hours).

The results of experimental on Twitter and Sina Weibo social media datasets show that KZWANG that we proposed model can significantly enhance the detection capability and be effectively discovered as early as possible at the same time.

6. Conclusions and Future Work

In this paper, we have proposed a rumor detection framework that combines text context semantic and propagate structural information to identify fake news and rumors. Different from most previous studies that use manually extracted features or deep learning automatic feature extraction methods, we fuse the contextual information from the source Weibo and corresponding reposts via multi-head attention, which generates a better semantic representation for each source Weibo. To capture the complex spread of information from different Weibo sources, we utilized a GCN to learn a heterogeneous graph constructed using propagate structural information to detect rumors. The extensive experiments conducted on real-world Weibo and Twitter datasets demonstrate the superiority of our proposed model compared to baseline models on rumor detection tasks.

Complex spread of information is an important feature of rumor, since constructing a heterogeneous graph which adequately represents propagation behavior is critical to the ability of GCN to learn propagation features. Graph neural network can better realize association mining

across spatial clues, Combining hidden clues in target data such as attributes, structures, and behaviors, mining complete clue information is still an important research direction. At the same time, how to introduce a large amount of language knowledge and world knowledge accumulated by human beings into the rumor detection model is an important direction to improve the performance of the fake news detection deep learning model, and it also faces significant challenges.

Author Contributions: Writing—original draft preparation, Z.K.; writing—review and editing, Z.L.; Data curation, C.Z. methodology, J.S.; supervision, W.S. Funding acquisition, Q.G.; Z.K. and Z.L. contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Xinjiang Uygur Autonomous Region Graduate Research and Innovation Project Grant No.XJ2020G071, Dark Web Intelligence Analysis and User Identification Technology Grant No.2017YFC0820702-3, the National Language Commission Research Project Grant No.ZDI135-96, and funded by China Academy of Electronics and Information Technology, National Engineering Laboratory for Risk Perception and Prevention (NEL-RPP).

Acknowledgments: We thank the anonymous reviewers for their valuable feedback.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Castillo, C.; Mendoza, M.; Poblete, B. Information credibility on twitter. In Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 28 March–1 April 2011; pp. 675–684.
- 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, Edinburgh, UK, 27–31 July 2011; pp. 1589–1599.
- Popat, K. Assessing the credibility of claims on the web. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, 3–7 April 2017; pp. 735–739.
- 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, Beijing China, 12–16 August 2012; pp. 1–7.
- Jin, F.; Dougherty, E.; Saraf, P.; Cao, Y.; Ramakrishnan, N. Epidemiological modeling of news and rumors on twitter. In Proceedings of the 7th Workshop on Social Network Mining and Analysis, Chicago, IL, USA, 11 August 2013; pp. 1–9.
- 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, Indianapolis, IN, USA, 24–28 October 2016; pp. 2377–2382.
- Ma, J.; Gao, W.; Wong, K.F. Detect rumors in microblog posts using propagation structure via kernel learning. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 708–717.
- Ma, J.; Gao, W.; Wong, K.F. Rumor detection on twitter with tree-structured recursive neural networks. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; pp. 1980–1989. [[CrossRef](#)]
- 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, Florence, Italy, 28 July–2 August 2019; pp. 1173–1179.
- Zhang, Q.; Lipani, A.; Liang, S.; Yilmaz, E. Reply-aided detection of misinformation via bayesian deep learning. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 2333–2343.
- Yu, F.; Liu, Q.; Wu, S.; Wang, L.; Tan, T. A convolutional approach for misinformation identification. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), Melbourne, Australia, 19–25 August 2017.
- Yu, F.; Liu, Q.; Wu, S.; Wang, L.; Tan, T. Attention-based convolutional approach for misinformation identification from massive and noisy microblog posts. *Comput. Secur.* **2019**, *83*, 106–121. [[CrossRef](#)]
- Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral networks and locally connected networks on graphs. In Proceedings of the International Conference on Learning Representations (ICLR2014), Banff, AB, Canada, 14–16 April 2014.

14. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. In Proceedings of the International Conference on Learning Representations, San Juan, PR, USA, 2–4 May 2016.
15. Zubiaga, A.; Aker, A.; Bontcheva, K.; Liakata, M.; Procter, R. Detection and resolution of rumours in social media: A survey. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–36. [[CrossRef](#)]
16. Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B.J.; Wong, K.F.; Cha, M. Detecting rumors from microblogs with recurrent neural networks. In Proceedings of the 25th International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9–15 July 2016; pp. 3818–3824.
17. Gao, J.; Han, S.; Song, X.; Ciravegna, F. RP-DNN: A tweet level propagation context based deep neural networks for early rumor detection in social media. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; pp. 6094–6105.
18. 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, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 549–556.
19. Lu, Y.J.; Li, C.T. GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Seattle, DC, USA, 5–10 July 2020; pp. 505–514. [[CrossRef](#)]
20. Yuan, C.; Ma, Q.; Zhou, W.; Han, J.; Hu, S. Jointly embedding the local and global relations of heterogeneous graph for rumor detection. In Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM), Beijing, China, 8–11 November 2019; pp. 796–805.
21. Ma, J.; Gao, W.; Joty, S.; Wong, K.F. An Attention-based Rumor Detection Model with Tree-structured Recursive Neural Networks. *ACM Trans. Intell. Syst. Technol. (TIST)* **2020**, *11*, 1–28. [[CrossRef](#)]
22. Wu, K.; Yang, S.; Zhu, K.Q. False rumors detection on sina weibo by propagation structures. In Proceedings of the 2015 IEEE 31st International Conference on Data Engineering, Seoul, Korea, 13–16 April 2015; pp. 651–662.
23. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
24. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 3844–3852.
25. Ma, J.; Gao, W.; Wei, Z.; Lu, Y.; Wong, K.F. 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, Melbourne, Australia, 19–23 October 2015; pp. 1751–1754.
26. Zhao, Z.; Resnick, P.; Mei, Q. Enquiring minds: Early detection of rumors in social media from enquiry posts. In Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 18–22 May 2015; pp. 1395–1405.
27. Kwon, S.; Cha, M.; Jung, K. Rumor detection over varying time windows. *PLoS ONE* **2017**, *12*, e0168344. [[CrossRef](#)] [[PubMed](#)]
28. Ma, J.; Gao, W.; Wong, K.F. Rumor detection on twitter with tree-structured recursive neural networks. In Proceedings of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018.
29. Liu, Y.; Wu, Y.F.B. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
30. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).