False Information Detection on Social Media

JINGHAN XU, College of Software, China

The paper designs a deep learning-based rumor detection method. The method, on the one hand, utilizes both the text information of the message and the message topology information, by integrating the social network message itself, the text information of the replies and the propagation topology information of the message to obtain the feature vectors of the message, and ultimately seeks for the optimal solution of the model by updating the parameters of the model through iterative updating; on the other hand, it applies the newly proposed GATv2 graph neural network model to deal with the topology information of the graph, which can better express the graph topology features through a dynamic attention mechanism in comparison with the conventional graph neural network, it can better express the graph topological features through the dynamic attention mechanism, so that the overall expression of the message is more representative. Using this idea can effectively improve the efficiency of rumor detection. Through experiments, the constructed model achieves 94.1%, 84.5% and 87.8% accuracy on three datasets, including Sina Weibo and Twitter, respectively.

CCS Concepts: • **Networks**; • **Network services**; • **In-network processing**; • **Network dynamics**;

Additional Key Words and Phrases: Rumor detection, deep learning, graph neural networks, natural language processing

ACM Reference Format:
Jinghan Xu. 2024. False Information Detection on Social Media. *J. ACM* 37, 4, Article 111 (June 2024), 8 pages. https://doi.org/1234567.89101123

1 INTRODUCTION

With the continuous development of communication technology, people's daily social life has truly realized the vision of far-flung realms, social network media has become the main means for people to obtain external information, every moment there are a huge amount of news in the social network dissemination and diffusion, in the face of the huge digital flood, which will inevitably be mixed with a large number of rumors. Chen Yanfang [7] and others define network rumors as "the interpretation or elaboration that social networks have played a key role in a certain stage or the whole process of their generation, dissemination and influence, and the content is not confirmed and has caused a certain impact on social opinion". When rumors appear in the audience's cognitive blind spot, they can easily be treated as real news, affecting the cognition and behavior of some users and being forwarded by some recipients to further spread the impact of misinformation. The influence of rumors is not to be underestimated, which may affect users' cognition and behavior for a period of time, or seriously threaten the stability of the society. For example, during the global spread of the New Crown Epidemic, the rumor "There is a microchip in the vaccine, which can control your mind" released by a Muslim cleric was wildly reposted in different languages on Facebook and Twitter, and the rumor was widely spread. This rumor was wildly reposted in different languages on Facebook and Twitter, and widely shared in many countries, causing confusion among the public and undermining citizens' trust in vaccination,

Author's address: Jinghan Xu, College of Software, Nan Jing, China, 1023041006@njupt.edu.cn.

ACM 0004-5411/2024/6-ART111
https://doi.org/1234567.89101123

showing the vulnerability of social media to rumors by this incident [10]. The study of automated monitoring of network false news has a certain practicality, the current methods for social network false news monitoring, according to the different kinds of features can be divided into two main types, based on text information features and based on the topological information between the social network messages, i.e., the propagation characteristics of the message.

1.1 Based on textual content

Jyoti et al [6] used Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) network as deep learning models for automatic feature extraction of tweeted text, and used Particle Swarm Optimization (PSO) algorithm to find the best features from the created hybrid feature set. Xiangwen Liao [12] et al. proposed the use of hierarchical attention network to segment the network time period and use GRU network for encoding to get the hidden representation of the time series. Kan Liu et al. [5] utilized the idea of transfer learning to complete the domain migration of the labeled information by incorporating the domain distribution variance based on the multi-head bi-directional LSTM model. Ma et al. [4] and Yu et al. [3] utilized recurrent neural networks and convolutional neural networks to learn message representations from time-series based message contents, respectively.

1.2 Based on propagation features

Rumor detection methods based only on text classification are mainly based on the mood or style of the news content, but if the false news is intentionally fabricated to mislead users, it is difficult to identify these finely processed false information with methods based only on text content. Therefore, in recent years, many scholars have focused on studying the role of social network message topology in improving the representational power of message feature vectors. To mislead a propagation pattern-based model, it is necessary to disrupt the overall propagation pattern including many ordinary users, which can provide stronger robustness compared to using text and user attributes that can be faked. To simultaneously extract useful information from post semantics and propagation structure, Wu et al [1] proposed a hybrid SVM classifier model to capture text patterns and propagation patterns in a sample set of Weibo posts. Liu et al [11] modeled the rumor propagation sequence as a multivariate time series and applied both recurrent and convolutional networks to capture user features along the rumor propagation path. Ma et al [8] used a tree kernel to capture similarities between propagation trees as a way to identify the propagation patterns of different types of rumors on the twitter social network platform. Ma et al [2] proposed the use of a tree-structured recurrent neural network to construct features based on post textual content features and message propagation topology. Lin et al [9] 2021 proposed the use of textual information fused with topological information for the solution of rumor detection problem, the GAT model is used in the proposed method to deal with the connection relationship between the original post of the published message and the forwarded post.

2 DATASET

2.1 Dataset description

In order to make the model generalizable, two datasets (Twitter and Weibo) are chosen, in English and Chinese, which contain textual information of false messages, reply information of each message, and topological information among all false messages, where the textual information of false messages is indexed by a unique ID, and the reply information of each message corresponds to a txt file, where each line corresponds to a reply, and the topological information is stored in a txt file, where each line corresponds to all the relevant neighbors of a message node. The data

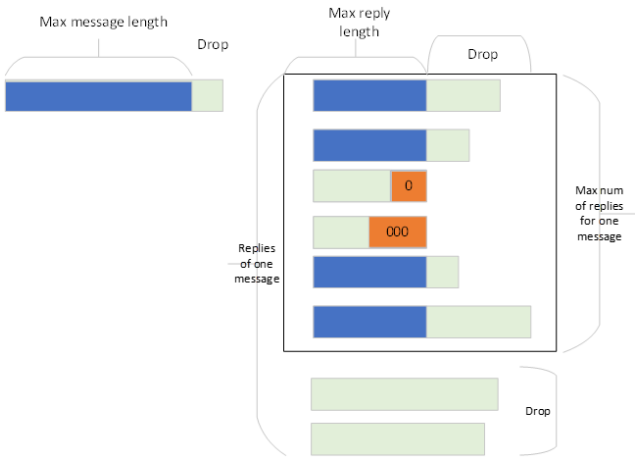

Fig. 1. Sample Vector Specification

generated by data preprocessing will be directly used as input to the model, so the method of data preprocessing will affect the algorithm model's learning of features to a certain extent. The data preprocessing is mainly divided into three parts: data cleaning, vectorization, and dataset slicing.

2.1.1 Data Cleaning. The dataset obtained from the experiment is usually obtained through the structure provided by the social network platform, and there is more noise in the dataset, which will affect the validity of the dataset and further affect the performance of the model if it is not processed. Therefore, data cleaning of the dataset is necessary. The specific approach to data cleaning in this paper is to convert the ambiguous escape characters present in the text content into strings, remove the unreadable emoticons, and delete the URLs present in the article.

2.1.2 Vectorization. Quantization that is, the text of the word embedded vectorization, this paper takes the word2vec method, specific practices:

a) *Create a dictionary*. Put all the message sentence text into a list, use the Counter method provided by python to count the occurrence of word frequency, to obtain the word frequency dictionary, where key is the text of the word, value is the word frequency, use the most common method of Counter to sort the words according to the word frequency, and use the sorted word number and the word as the key and the value to form the dictionary, respectively, where remove the word frequency of 1, and use the most common method as the key and value. Among them, the words with word frequency of 1 are removed, because the frequency of occurrence is too low, indicating that it is a rare word, which has little effect on the generation of semantics.
b) *Sentence indexing*. The words appearing in the sentence are transformed into the value corresponding to the word in the dictionary to obtain the serialized sentence. In order to data uniformity and facilitate the calculation of the subsequent model, set the fixed sentence length l1 of the original post, the maximum number of reposts l2 of a single original post and the maximum sentence length l3 of reposts, for the indexed sentences whose sentence lengths are lower than the limited length, use index 0 instead of the complement, for the indexed sentences whose sentence lengths are more than the limited length, truncate, and for the number of replies less than the limited number of replies to a single original post, use

the all-zero sentence Completion, for the original post with more than the limited number of replies, the same truncation, the above specification process is shown in Figure 3.4.

c) *Indexing vectorization.* Iterate through all the text sentences, use the Embedding method provided by the nn library in pytorch, where the weights use the pre-trained word2vec model weights, and set the padding attribute to 0, i.e., specify that the elements with an index of 0 are extended bits and do not have semantics when vectorizing. Finally obtain the vectorization results of all the text.

2.1.3 Dataset slicing. . For the training and prediction of the model, the dataset is randomly sliced by 9:1, and then the data are randomly extracted from the training set as the validation set with a ratio of 0.2. The role of the training set is to update the parameters and find the optimal model parameters; the role of the validation set is to test the training results of the training set periodically and adjust the model according to the results; and the role of the test set is to check the final training results.

3 FALSE INFORMATION DETECTION

The overall flow of the method is shown in Fig.2. It can be roughly divided into the following steps:

(1) Coding the fake messages as well as replies for text feature extraction respectively
(2) Fusing the message itself (dimension) and the feature vector obtained from the reply (dimension) into a single feature vector, characterized to represent the textual features of the message
(3) Input the topological information into the graph neural network model GATv2, and the resulting data are the topological feature vectors corresponding to each message
(4) Fuse the textual feature vectors and topological feature vectors of the messages (i.e., splicing and then dimensionality reduction) to obtain the fused representation vector of the final message.

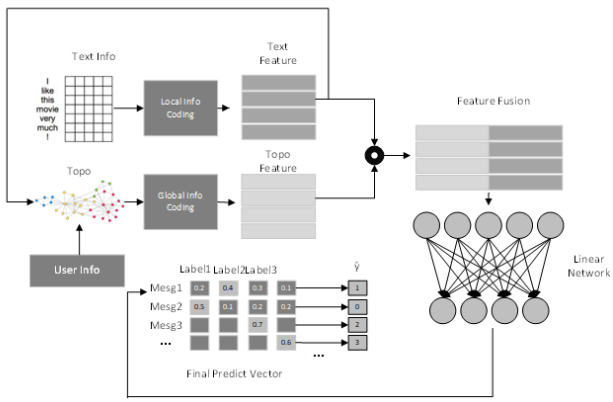

Fig. 2. Overall framework

Here the text feature extraction is explained in detail, text feature extraction is to extract features from the text representation vectors obtained in the preprocessing stage, i.e., through certain transformations, it is transformed into feature vectors with lower dimensionality and stronger representativeness, and in order to better integrate the contextual information, we use CNN in combination with LSTM to process the text vectors.

In order to rationally fuse source messages with fused reply comment messages, a fusion gate structure is designed:

$$\alpha = \sigma\left(w_1 m + w_2 r + b\right) \tag{1}$$

$$\widetilde{M} = m \bigodot \alpha + r \bigodot (1 - \alpha) \tag{2}$$

where σ is the sigmoid activation function, w and b are the learnable parameters, and $\widetilde{M}$ is the final message representation. In many cases, the importance between a message and its forwarded comments is not equally important, sometimes the message appearance provides more information for the determination, which can provide enough basis for the determination, and sometimes the finely processed message has a more confusing appearance, while its forwarded comments provide more information. For such a scenario of changing importance, if the conventional splicing, or other methods that default to equal importance of the two, are used, then it will result in a certain degree of information loss, while the gate structure given in this paper can provide the model with greater flexibility to make up for the above shortcomings.

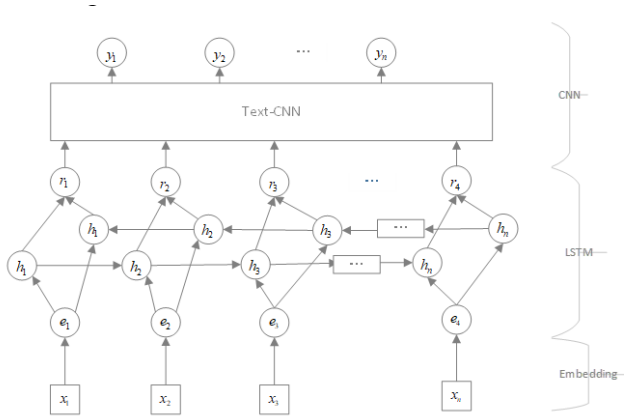

Fig. 3. Text Processing Framework

Firstly, the initial vectors are fed into the LSTM neural network, so that each bit in the vector has temporal information, and then the output vectors are used as inputs to the CNN, so that the feature vectors have contextual information, in which the CNN is set to three layers, the convolution kernels are 3,4,5, and the corresponding pooling layers are also set to three layers, and the final outputs are the feature vectors of the text.

Since the graph neural network can be directly imported by python's package, and since the dimensions of the input and output need to correspond to the provided data and the length of the text vectors, there are fewer parameters to be adjusted, but it is found through experiments that the difference in the number of its layers has a certain impact on the prediction accuracy of the model. As shown in Table:

Table 1. EFFECT OF NUMBER OF LAYERS ON RESULTS

Layers	Precision
1	0.941
2	0.922
3	0.910

It can be seen that, similar to other neural network properties, it is not the case that the more the number of layers the better the training results obtained, because as the number of layers increases, the function complexity of the model increases and the representational power is better, but at the same time, its probability of overfitting increases considerably, and even though there may be a high level of accuracy on the training set, it may be less accurate on the test set.

4 RESULT

By validating on two different datasets[7], it can be found that the model on the English dataset achieves a similar performance to the benchmark model, and the accuracy is more stable, while on the Chinese dataset the difference is more obvious, and the model has a large advantage, which has a tendency to increase with the increase in the number of iterations.

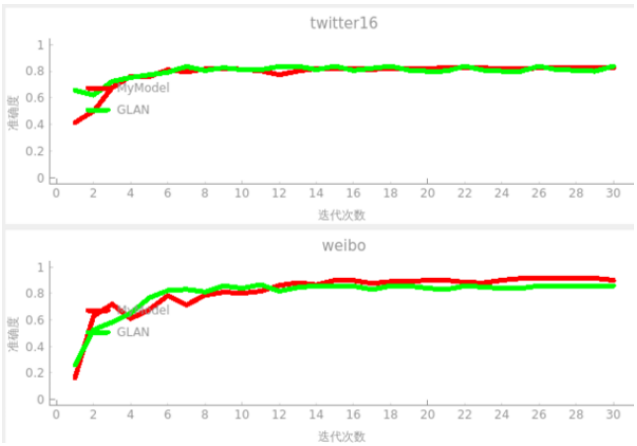

Fig. 4. Comparison of experimental results

A batch of samples is selected from the sample set to simulate real social network messages for prediction, and the prediction results are visualized, and it can be found that the model has a certain degree of accuracy for screening false messages.

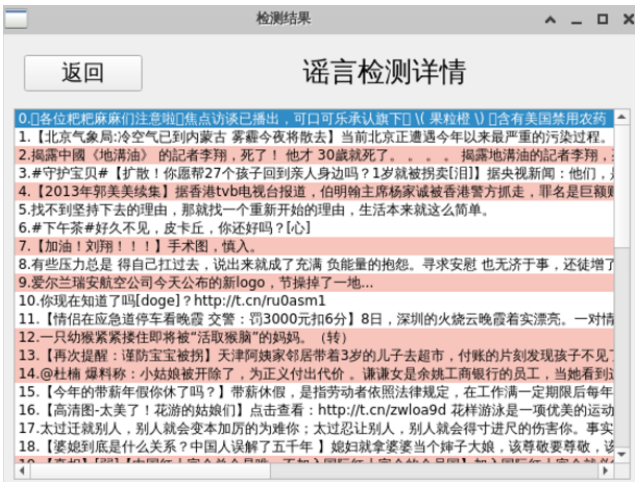

Fig. 5. Visualization of results

5 CONCLUSIONS

The problem of rumor intelligent detection will always be accompanied by the development of social networks, which is the core problem to ensure that the public can avoid the influence of rumors to the greatest extent and maintain social stability. In this paper, we constructed a rumor intelligent detection system based on deep learning by studying and learning the existing theoretical results, based on the personal understanding of the problem, and fusing the advantages of the text extraction model and the graph neural network model, and made some results on the rumor detection problem.

The main research work of the thesis is as follows:

(1) A general overview of the recent research on the rumor detection problem at home and abroad, a detailed study of the text extraction, graph neural network and other models, and based on the reflection and online access to give the significance of the current direction of this research and the solution of the rumor detection problem. And the excellent algorithms of the predecessors are borrowed and studied and theoretically analyzed.
(2) Modeling the social network and giving the relationship structure among original post, repost and user.
(3) The latest graph neural network model GATv2 is applied to the field of rumor detection, and experimentally verified that the model has certain performance superiority in this field, and can obtain the feature vectors of the nodes in the topological graph which are more representative.
(4) In the process of text vectorization, the advantages of Text-CNN and Text-RNN are integrated, and global information is added to the vectors obtained by convolutional neural network with limited field of view, which improves the representativeness of text features.
(5) Using the gate mechanism, the original post features are embedded into the repost feature vector to ensure that the repost will not deviate from the topic. In the final part of the fusion of global features and local features, the gate mechanism is also used, so that the final fusion features retain a certain degree of elasticity in their attention to the information of the two parts.

The following conclusion is drawn from the above study: the intelligent rumor detection mechanism based on deep learning is feasible and effective, and improves the efficiency of similar as-improved algorithms to a certain extent.

REFERENCES

[1] 1984. *SIGCOMM Comput. Commun. Rev.* 13-14, 5-1 (1984).

[2] Sem Borst, Varun Gupta, and Anwar Walid. 2010. Distributed Caching Algorithms for Content Distribution Networks. In *2010 Proceedings IEEE INFOCOM*. IEEE, San Diego, CA, USA, 1–9. https://doi.org/10.1109/INFCOM.2010.5461964

[3] Mauro Conti, Roberto Di Pietro, Luigi V. Mancini, and Alessandro Mei. 2009. (old) Distributed data source verification in wireless sensor networks. *Inf. Fusion* 10, 4 (2009), 342–353. https://doi.org/10.1016/j.inffus.2009.01.002

[4] Honghao Gao, Wangyang Jiang, Qionghuizi Ran, and Ye Wang. 2024. Vision-Language Interaction via Contrastive Learning for Surface Anomaly Detection in Consumer Electronics Manufacturing. *IEEE Transactions on Consumer Electronics* (2024), 1–12. https://doi.org/10.1109/TCE.2024.3378771

[5] Honghao Gao, Binyang Qiu, Ye Wang, Si Yu, Yueshen Xu, and Xinheng Wang. 2023. TBDB: Token Bucket-Based Dynamic Batching for Resource Scheduling Supporting Neural Network Inference in Intelligent Consumer Electronics. *IEEE Transactions on Consumer Electronics* (2023), 1–12. https://doi.org/10.1109/TCE.2023.3339633

[6] Honghao Gao, Xuejie Wang, Wei Wei, Anwer Al-Dulaimi, and Yueshen Xu. 2024. Com-DDPG: Task Offloading Based on Multiagent Reinforcement Learning for Information-Communication-Enhanced Mobile Edge Computing in the Internet of Vehicles. *IEEE Transactions on Vehicular Technology* 73, 1 (2024), 348–361. https://doi.org/10.1109/TVT.2023.3309321

[7] Haojia He, Songtao Guo, Lu Yang, and Ying Wang. 2022. MACC: MEC-Assisted Collaborative Caching for Adaptive Bitrate Videos in Dense Cell Networks. In *2022 18th International Conference on Mobility, Sensing and Networking (MSN)*. 218–222. https://doi.org/10.1109/MSN57253.2022.00046

[8] IEEE 2004. IEEE TCSC Executive Committee. In *Proceedings of the IEEE International Conference on Web Services (ICWS '04)*. IEEE Computer Society, Washington, DC, USA, 21–22. https://doi.org/10.1109/ICWS.2004.64

[9] IEEE 2004. IEEE TCSC Executive Committee. In *Proceedings of the IEEE International Conference on Web Services (ICWS '04)*. IEEE Computer Society, Washington, DC, USA, 21–22. https://doi.org/10.1109/ICWS.2004.64

[10] Yaqi Song and Shenyun. 2023. Video Stream Caching Based on Digital Twin Cooperative Caching. In *2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '23)*. IEEE, Beijing, China, 1–5. https://doi.org/10.1109/BMSB58369.2023.10211139

[11] Renato Werneck, João Setubal, and Arlindo da Conceicão. 2000. (old) Finding minimum congestion spanning trees. *J. Exp. Algorithmics* 5 (2000), 11. https://doi.org/10.1145/351827.384253

[12] Jiaxin Xu, Huiling Shi, Haoxiang Chu, and Wei Zhang. 2023. EeCA: A Novel Approach for Energy Conservation in MEC via NDN-Based Content Caching. In *2023 24st Asia-Pacific Network Operations and Management Symposium (APNOMS)*. 349–352.

Received 28 May 2024; revised 20 June 2024; accepted 28 June 2024