Postprint of Rumor Identification on Weibo in Public Health Emergencies

Authors: Shi Kaiwen, Liu Kan

Date: 2023-04-01T16:02:52+00:00

Abstract

[Purpose/Significance] In public health emergencies such as the "COVID-19" pandemic, a large volume of pandemic-related content is rapidly generated on social media platforms, which includes a considerable amount of deliberately spread rumors. These not only endanger public mental health but also affect the implementation of response strategies for public health emergencies. Identifying rumors in public health emergencies enables the public to face crises correctly and plays a positive role in maintaining social stability and cyberspace governance. [Method/Process] First, we conduct an in-depth analysis of collected rumors that have been verified during the pandemic period to extract the main features of rumor texts, including contextual features, topic category features, sentiment intensity features, keyword features, etc. Then, to address the issue of relatively single-dimensional text feature representation in text classification models, we utilize different models to vectorize the extracted rumor text features and enhance and fuse various types of text features. Among them, the word vector weights computed through TF-IDF can strengthen keyword feature information at the word granularity while capturing contextual features. Finally, we employ a BiLSTM+DNN model to classify and discriminate the fused feature vectors. [Results/Conclusion] Experimental results demonstrate that features such as topic category and sentiment intensity all contribute to rumor identification; particularly, the fusion of enhanced word vectors with other features leads to significant improvement in identification accuracy, with metrics such as recall and F1-score all exceeding 90%, outperforming other rumor identification models. This indicates that the proposed method can effectively achieve rumor identification in the context of public health emergencies.

Full Text

Preamble

Weibo Rumor Identification in Public Health Emergencies

Shi Kaiwen, Liu Kan
School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073

Abstract:
[Purpose/Significance] In public health emergencies such as the COVID-19 pandemic, social media platforms rapidly generate massive amounts of epidemic-related discourse, including numerous intentionally spread rumors that not only endanger public mental health but also impede the implementation of public health response strategies. Identifying rumors in such contexts enables citizens to face crises correctly and plays a positive role in maintaining social stability and network governance. [Method/Process] We first conducted an in-depth analysis of confirmed rumors collected during the pandemic, extracting primary features of rumor texts including contextual features, topic category features, sentiment intensity features, and keyword features. To address the problem of relatively single-dimensional text feature representation in text classification models, we utilized different models to vectorize the extracted rumor text features and performed enhancement and fusion of various text features. Notably, word vector weights calculated through TF-IDF can capture contextual features while strengthening keyword feature information at the word granularity. Finally, a BiLSTM+DNN model was employed to classify the fused feature vectors. [Result/Conclusion] Experimental results demonstrate that features such as topic category and sentiment intensity all contribute to rumor identification. In particular, the fusion of reinforced word vectors with other features significantly improves identification accuracy, with recall, F1-score, and other metrics all exceeding 90%, surpassing other rumor identification models. This indicates that our proposed method effectively achieves rumor identification in the context of public health emergencies.

Keywords: public health emergencies; rumor identification; Weibo; multi-feature fusion
Classification Number: TP393
DOI: 10.13266/j.issn.0252-3116.2021.13.009

---

1. Introduction

According to the *Regulations on Public Health Emergency Response* [1], public health emergencies are defined as sudden occurrences that cause or may cause serious harm to public health, including major infectious disease outbreaks, unexplained diseases in groups, major food and occupational poisoning, and other events that seriously affect public health. Such emergencies are closely related to public life and simultaneously trigger large volumes of online rumors on social media platforms like Weibo and WeChat. Weibo plays a crucial role in information dissemination and public opinion guidance during public health events, but also serves as a breeding ground and channel for rumor proliferation. Rumors on Weibo have greater exposure than on WeChat, forums, and other online media, with wider dissemination, stronger deceptive capacity, and deeper impact. Therefore, we focus our research scope on rumors on the Weibo platform, where extracting effective features—especially content features—from public health event rumors plays a critical role in identification [2].

Regarding feature extraction for Weibo, different research perspectives emphasize different features, with most current research concentrating on content features and user characteristics. Content features focus on contextual, semantic, and multimedia features of Weibo posts, while user features examine behavioral and influence characteristics. Due to the unique nature of public health emergencies, we concentrate on content features of Weibo texts, analyzing differences between texts during public health emergencies and normal times, using various text feature representation methods to analyze and extract text features from multiple perspectives, then fusing these features to construct an effective Weibo rumor detection model for public health emergencies.

Rumors about public health emergencies pose significant threats to social stability and people's livelihood security, creating major obstacles for public psychological stability and government governance. For instance, during the early stages of COVID-19, the online rumor that "only N95 masks are effective for virus prevention" misled the public into panic-buying and hoarding N95 masks, severely impacting normal virus prevention efforts. Therefore, rumor identification in public health emergencies is urgent and important. However, such rumors are particularly difficult to identify due to their highly deceptive nature, strong emotional content, and high public attention.

Weibo is one of China's most widespread platforms for discourse dissemination. In public health emergencies, it serves as both an information channel and a rumor vector. Existing research on rumor identification has primarily focused on conventional scenarios, with features often requiring specialized design by researchers. Deep learning-based rumor identification requires large training datasets. For rumor identification in specific domains like public health emergencies, it is necessary to research and extract domain-specific features such as topic content, linguistic characteristics, and emotional polarity. Meanwhile, fusion of different features also plays a significant role in rumor identification. Consequently, we design and extract multiple content features from rumor texts, explore effective fusion methods, and construct a rumor identification model tailored to specific public health emergencies.

2. Related Work

Research on rumor identification has primarily proceeded along two dimensions: feature extraction from rumors and algorithm design for identification. Most studies focus on extracting effective features from rumor data, while others concentrate on identification algorithms.

In terms of rumor feature research, the earliest work by C. Castillo et al. [3] evaluated news credibility on Twitter, extracting 68 features covering message basic features, user features, topic features, and propagation features. Subsequently, F. Yang et al. [4] used Weibo rumor data to propose two new features: client software used by users and geographical location of events. He Gang et al. [5] introduced a series of new features including symbol features, link features, word frequency distribution features, and time differences, combining them with Weibo text and user features. Xia Song et al. [6] constructed a sensitive word lexicon through a newly designed word extraction algorithm, adding sensitive word features to basic features like content and user behavior, significantly improving rumor identification accuracy. Li Gang et al. [7] proposed a novel rumor propagation model for coupled social networks based on audience profiling, constructing multi-dimensional functions from basic features including audience cognitive ability, anonymity, and authority, as well as psychological features such as conformity, memory effects, and friend influence.

Temporal features of rumors have also received considerable attention. S. Kwon et al. [8] first highlighted the importance of temporal attributes in rumor propagation, studying propagation characteristics across time, structure, and language to determine rumor features, then constructing a random forest classifier. J. Ma et al. [9] extended this work by further expanding the time-varying feature set, using simple equal-interval time series segmentation to observe how rumor event features change over time—a modeling technique applied to integrate various social context information. Wang Zhihong et al. [10] introduced fuzzy time series model domain partition concepts to better observe and represent temporal changes in rumor event features, proposing a dynamic partitioning algorithm for event temporal data based on fuzzy clustering to construct time-varying feature sets. M. Kotteti et al. [11] proposed a multi-time-series data analysis model for rumor detection on Twitter that uses only temporal characteristics of tweets instead of content, significantly reducing computational complexity for rapid rumor detection.

Regarding rumor identification algorithms, besides traditional machine learning techniques like logistic regression, SVM, and random forest, increasing numbers of researchers employ deep learning models. J. Ma et al. [12] leveraged recurrent neural networks' advantage in context extraction to analyze connections between posts on a topic, reducing the impact of low-value posts on rumor identification. L. Li et al. [13] proposed a rumor detection method based on deep bidirectional gated recurrent units (D-Bi-GRU) to capture forward and backward contextual features of Weibo streams, obtaining group response information as

Weibo events evolve. Wang Xingyu [14] constructed three deep features added to basic shallow features, including sentence vectors built by Doc2vec and sentiment polarization calculated by the SnowNLP sentiment analysis library. Liu Kan et al. [15] applied transfer learning to Twitter rumor detection. G. Siva et al. [16] constructed graph features for fake news, performing unsupervised learning based on graph feature vector learning and label propagation algorithms. F. Marra et al. [17] used adversarial generative networks to identify fake images in social networks, while F. Qian et al. [18] applied text generation technology to fake news identification, using generators to create comments about news events to judge authenticity.

However, current research on rumor identification in public emergencies or social hotspot events remains limited. Fan Rong et al. [19] selected rumor texts related to the 2016 "Shandong illegal vaccine incident" and "Mizhi No.3 Middle School injury incident" on Weibo, constructing an R-CNN identification model based on users' historical texts, rumor attention levels, and Weibo frequency. Zeng Ziming et al. [20] used LDA topic models and random forest algorithms to detect 2016 haze rumors. Wang Lin et al. [21] established an influencing factor model for public health emergency public opinion dissemination based on ELM, TAM models, and lifecycle theory, using feature variables such as information content, publication date, and publisher certification type. Li Lihua et al. [22] selected five terrorist attack incidents in the UK in 2017 as case studies, analyzing dissemination subjects, information content features, and propagation characteristics on Twitter datasets to study rumor propagation patterns.

Overall, feature extraction for rumor identification often relies on specialized researcher design, with most rumors focusing on conventional scenarios. Deep learning-based rumor identification requires large training datasets. For rumor identification in specific domains like public health emergencies, it is necessary to research and extract domain-specific features such as topic content, linguistic characteristics, and emotional polarity. Meanwhile, fusion of different features also significantly impacts rumor identification. Therefore, we design and extract multiple content features from rumor texts, explore effective fusion methods, and construct a rumor identification model tailored to specific public health emergencies.

---

3. Weibo Rumor Feature Analysis

3.1 Rumor Data Sources

Using the COVID-19 pandemic as an example of a public health emergency, we collected confirmed rumor data from Weibo Community Management Center's official rumor-refutation account. The timeframe covered the first four months of concentrated rumor activity since the outbreak, from January 1 to April 30, 2020, totaling 730 Weibo posts. Since many Weibo messages appear unreliable but few are actually verified and refuted—demonstrating that manual rumor refutation is extremely difficult and time-consuming—available annotated data is limited, restricting the use of deep learning models. Therefore, our research focuses on rumor feature extraction and fusion. For comparison and data balance, we additionally collected 1,400 non-rumor Weibo posts during the pandemic from official accounts such as Xinhua News Agency, People's Daily, and CCTV to ensure authenticity. Data collection used Selenium to simulate browser access, BeautifulSoup to parse webpage content, and regular expressions with find functions to match required fields. Figure 1 [Figure 1: see original paper] shows sample collected data.

3.2 Pandemic Rumor Text Analysis

Compared with general online public opinion events, public health emergencies exhibit not only general characteristics like suddenness, directness, interactivity, and immediacy, but also unique features including high public participation, strong negative tendencies, intensified public panic, and extensive use of extreme language. Weibo text features manifest primarily in the following aspects:

(1) Rumor Word Cloud. Based on rumor texts, we generated the word cloud shown in Figure 2 [Figure 2: see original paper], where five terms are most prominent: "infection," "prevention," "mask," "treatment," and "school opening." When public health emergencies erupt, public attention focuses primarily on disease and livelihood issues. Information about life safety, food and clothing security, transportation, and school/work resumption becomes most sought-after. Rumor mongers exploit these concerns, releasing related rumors to attract attention, such as claims about COVID-19 prevention, treatment, virus transmission routes, and public transportation. These often follow fixed patterns like "…is…, please be informed" or "only…can…, remember." Such statements are typically alarmist, leveraging high public attention in specific contexts to shape public opinion.

(2) Word Frequency Features. In pandemic rumors, certain words appear frequently, including intensifiers like "extremely," "very," "must," and "absolutely," which create urgency. Signature terms related to public health emergencies such as "virus prevention," "prevention," "hospital," "expert," and "alcohol" appear repeatedly. For comparison, the right half of Table 1 shows high-frequency words in non-rumor texts, where more formal terms like "recently," "response," and "implementation" dominate, with exaggerated adverbs nearly absent.

(3) Keyword Features. To further analyze word usage patterns in rumor texts, we used TF-IDF (Term Frequency-Inverse Document Frequency) to calculate word weights across different rumors, reflecting keyword significance. Table 2 left shows keywords from pandemic rumors, with terms like China, USA, Wuhan, lockdown, medical care, hospital, pandemic, and pneumonia being most prominent. In contrast, non-rumor texts (Table 2 right) feature keywords like live broadcast, news, release, life, health, and information, focusing on daily information.

(4) Topic Features. We used LDA topic models to extract topic features from Weibo rumor texts. Using a topic vector structure similarity minimization method [23], we determined the optimal number of topics as seven. The extracted topic vectors are shown in Table 3 , covering seven major themes: disease prevention, disaster relief, policy interpretation, figure focus, livelihood security, pandemic dynamics, and popular science. Among these, disease prevention, disaster relief, and livelihood security accompany various public health events. However, rumors focusing on figures (e.g., "Zhong Nanshan: No vegetarians among confirmed cases," "Li Lanjuan says dating can prevent COVID-19"), policy misinterpretations (e.g., "XX location will also be locked down," "Government will use aircraft for disinfection spraying"), and pseudo-scientific claims (e.g., "Hair dryers can disinfect masks," "Viruses die at 20°C air conditioning") are closely tied to specific events like COVID-19 and represent topics of greatest public concern—precisely what rumor spreaders exploit to confuse the public.

(5) Sentiment Features. Rumor sentiments are often stronger and richer than normal discourse. Using the HowNet sentiment analysis lexicon, we analyzed sentiment in 730 rumor and 730 non-rumor samples, obtaining the sentiment score distribution shown in Figure 3 [Figure 3: see original paper]. Sentiment scores range from 0 to 1, with values near 1 indicating positive emotion and near 0 indicating negative emotion. The results show that hot events exhibit clear emotional polarization, concentrated at both ends of the distribution. Compared with normal discourse where positive sentiment dominates absolutely, rumors show a significantly increased proportion of negative sentiment, approaching the level of positive sentiment.

---

4. Model Construction

4.1 Basic Approach

Based on the Weibo rumor text features analyzed above for public health emergencies, we focus on utilizing topic category features, sentiment features, and keyword features for rumor identification (with word frequency features incorporated in keyword weight calculation). Features from different perspectives can complement each other. We then vectorize and fuse these features, including text enhancement of keyword features and concatenation with topic and sentiment features, before using deep learning networks for binary classification.

4.2 Model Process

The constructed rumor identification model is shown in Figure 4 [Figure 4: see original paper], comprising a basic feature extraction layer, feature fusion layer, and classification layer. The basic feature extraction layer first preprocesses input text through tokenization and stop-word removal, then calculates basic features using trained models: semantic word vector features from Word2Vec, keyword features (f1) from TF-IDF, topic features (f2) from LDA, and sentiment features (f3) from sentiment lexicons. The feature fusion layer includes: (1) strengthening word vectors using keyword weights to highlight important terms, and (2) concatenating all feature types into a final fused feature vector. The classification layer employs a Bidirectional Long Short-Term Memory (BiLSTM) and Deep Neural Network (DNN) to classify the concatenated feature vector and output class labels.

(1) Keyword Enhancement. Word vector features are crucial for text representation. A key step in our model is weighted enhancement of keyword vectors. Word vectors calculated using Word2Vec during pretraining effectively capture contextual features, but lack emphasis, making it difficult to reflect core terms in statements. Incorporating TF-IDF weights strengthens keyword vector weights, making word vectors more focused and better representing text features. Figure 5 [Figure 5: see original paper] illustrates the TF-IDF weight strengthening process.

In Figure 5, let D represent the document collection, with di denoting a tokenized and stop-word-removed document. Let W represent the vocabulary, with wi denoting a word. Equation (1) shows the vectorization of each document:

$$\mathrm{docVec}(d_i) = \frac{1}{\mathrm{length}(d_i)} \sum_{w_i \in d_i} \mathrm{tfidf}_{d_i}(w_i) \cdot \mathrm{WordVec}(w_i) \times C$$

where length(di) represents the number of words in the document, C is a weight adjustment coefficient to avoid gradient vanishing, and tfidf_{di}(wi) represents the TF-IDF weight of the word in the document. WordVec(wi) represents the vector of word wi in corpus D. This formula essentially broadcasts the TF-IDF weight into the word vector, enabling the vector to incorporate both contextual and keyword information. Through TF-IDF weight broadcasting, the same word gains different vector representations in different documents, enriching and completing its embedded meaning.

(2) Feature Concatenation. After fusing Word2Vec with TF-IDF keyword weights, we obtain new word vectors f1' for each word in rumors. Using the LDA model, we calculate the probability distribution of each rumor across seven topics, yielding a 7-dimensional topic distribution vector f2. Sentiment calculation uses SnowNLP scores normalized via z-score to transform sentiment scores to a -1 to 1 range, obtaining sample sentiment features f3. Finally, we concatenate these feature vectors (f1' + f2 + f3) and feed them into the classification model.

(3) Rumor Discrimination Model. After feature fusion, we construct a classifier for Weibo rumor identification. We use a BiLSTM+DNN network: the BiLSTM layer extracts features, concatenates forward and backward sequential outputs, then feeds them into DNN and output layers to predict whether samples are rumors. The specific structure is shown in Figure 6 [Figure 6: see original paper].

Concatenating features f1', f2, and f3 yields the fused feature vector Xf, where Xf = [f1', f2, f3] = {X1^f, ..., Xi^f, ..., Xn^f}, with vector dimension n = |f1'| + |f2| + |f3|. Inputting Xf into the BiLSTM layer produces forward output Lh = {Lh1, ..., Lhi, ..., Lhk} and backward output Rh = {Rh1, ..., Rhi, ..., Rhk}. Concatenating these yields H = [Lh, Rh], where |H| = 2k. The fully connected layer has weights Wh and bias bh, outputting H' = {H1', ..., Hi', ..., Hh'}. The output layer judges whether content is rumor, with weights wo and bias bo, outputting ŷ as shown in Equations (2) and (3):

$$H' = W_h H + b_h \quad (2)$$

$$\hat{y} = \sigma(W_o H' + b_o) \quad (3)$$

---

5. Experiments and Results

5.1 Experimental Setup

The experimental objective is rumor identification regarding COVID-19 on Weibo. We established four control groups to test different conditions and our proposed model's performance:

(1) Traditional machine learning methods: Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), and ensemble learning (XGBoost).

(2) Representative deep learning models: Given data limitations, besides basic Convolutional Neural Networks (CNN), we included Transfer Learning (TL) and Generative Adversarial Networks (GAN). TL used a BiLSTM network trained on historical rumor datasets from literature [16] to construct a pandemic rumor classifier. GAN artificially generated 3,000 new fake information samples, with tests showing optimal performance when adding 1,500 generated samples, indicating that balance between original and generated data affects results.

(3) Rumor identification without TF-IDF text enhancement: Using original pretrained word vectors f1 combined with topic features f2 and sentiment features f3 to compare enhancement effects.

(4) Rumor identification with TF-IDF enhancement (our model): Using enhanced word vectors f1' combined with topic features f2 and sentiment features f3.

5.2 Experimental Process

Using text feature vector f1 as input, we built classifiers with traditional machine learning models (NB, SVM, DT, XGBoost) and deep learning models (CNN, TL, GAN). We then added topic features f2 and sentiment features f3 to f1, building classifiers with the BiLSTM+DNN model. Finally, using enhanced text feature vector f1' as input with f2 and f3, we built classifiers with BiLSTM+DNN to compare with Group 3 and reflect word vector weight enhancement effects. The BiLSTM+DNN training loss is shown in Figure 7 [Figure 7: see original paper].

Data preprocessing used the Jieba tokenizer and Harbin Institute of Technology stopword list, with background terms like "pandemic" and "COVID" added to the segmentation dictionary. Word2Vec's CBOW algorithm vectorized texts into 300-dimensional vectors per word. To prevent overfitting and improve robustness, we applied weight decay for learning rate decay and dropout in LSTM layers. Main parameters are shown in Table 4 .

Table 4 Experimental Parameter Settings

Parameter	Value
Epoch	20
Dropout	0.5
Weight_{decay}	0.001
Batch_{size}	32
Classification layer activation function	sigmoid
BiLSTM hidden layer neurons	128
BiLSTM hidden layers	2
DNN fully connected layer neurons	64
Weight adjustment coefficient C	10
CBOW word vector dimension	300

Experiments ran on Deepin 20 OS with Python 3.8, using PyTorch to build and train the rumor identification model.

5.3 Experimental Results and Analysis

We used five-fold cross-validation, evaluating models with recall, precision, and F1-score, with recall being most critical—measuring the proportion of rumors that should be identified that are actually identified. Results are shown in Table 5 .

Table 5 Public Health Emergency Weibo Rumor Identification Results

Model	Recall	Precision	F1-score
Group 1: Traditional Machine Learning			
NB	0.5122	0.4627	0.4862
SVM	0.5478	0.5870	0.5667
DT	0.6630	0.7492	0.7034
XGBoost	0.7914	0.7215	0.7548
Group 2: Deep Learning Models			
CNN	0.7382	0.6365	0.6836
TL	0.8286	0.7781	0.8026
GAN	0.8618	0.8206	0.8407
Group 3: Without TF-IDF Enhancement			
f1+f2	0.8890	0.8081	0.8428
f1+f3	0.8657	0.9805	0.9181
f1+f2+f3	0.8972	0.8495	0.8695
Group 4: With TF-IDF Enhancement (Our Model)			
f1'+f2	0.8958	0.9622	0.9271
f1'+f3	0.9328	0.8586	0.8913
f1'+f2+f3 (Our Model)	0.9109	0.9363	0.9199
f1'+f2+f3 (Full)	0.9602	0.9709	0.9654

Our methods demonstrate clear advantages. In Group 1, traditional machine learning models show large recall variations, with ensemble learning performing best but still below 80%, limited by both dataset size and model simplicity. In Group 2, deep learning models improve recall, but CNN underperforms traditional methods due to lack of data and feature advantages. Transfer learning achieves only 83% recall because general rumor data and pandemic rumor data differ significantly in text features, language distribution, and domain. GAN performs best, showing that increased data volume improves identification, though excessive generated data breaks balance with original data and reduces effectiveness.

Group 3 clearly shows that adding topic and sentiment features substantially improves model performance. Topic features (f2) notably improve precision and F1-score, while sentiment features (f3) improve recall. Models fusing both features demonstrate overall advantages in recall, precision, and F1-score, confirming feature complementarity.

Group 4 shows that after word vector weight enhancement and fusion with other features, our model achieves the best recall and F1-score, with precision ranking second. Most importantly, recall—the critical metric for rumor identification—improves significantly compared to other methods. Most experiments also show clear F1-score improvements, far surpassing original word vector models. This demonstrates that word vector strengthening creates better complementarity with other features, achieving superior results.

We also compared our approach with existing rumor-refutation platforms like

China Internet Joint Rumor Refutation Platform, Weibo Rumor Refutation, and Science Rumor Refutation Network, which rely on manual verification without search functionality, failing to enable early rumor detection and containment. Tencent News's pandemic rumor verification platform "Jiaozhen" matches user inputs with official news but lacks deep algorithmic support, unable to accurately identify many rumors that our model correctly classifies, demonstrating our model's practical utility.

---

6. Conclusion

Weibo rumor identification in public health emergencies plays a vital role in maintaining network and social stability. Research in this context provides positive significance for future rumor identification in other emergencies. Due to unique event backgrounds, traditional rumor identification methods have limited effectiveness, and general deep learning models struggle with limited data volumes, making text feature extraction critical. Building upon semantic word vector training, we added topic category features, sentiment features, and keyword features, designing a TF-IDF-based word vector strengthening method. After merging with other features, we constructed a BiLSTM+DNN-based rumor identification model. Experimental results show our model outperforms existing approaches, achieving excellent results.

Future work could incorporate additional user behavior features, temporal sequence features, propagation features, and publisher identity features, though current datasets for these features remain incomplete and involve overly complex processing with unproven effectiveness and efficiency. Additionally, our word embedding enhancement operates at word granularity; future research could explore enhancement methods at different granularities or with different preprocessing approaches.

References

[1] State Council Policy Website. Regulations on Public Health Emergency Response [EB/OL]. [2021-05-29]. http://www.gov.cn/zhengce/2020-12/26/content_{5574586}.htm.

[2] Zhang Li, Mao Haoran. Review of Research on Online Rumor Governance in Public Emergencies [J]. Journal of Nanjing Xiaozhuang University, 2020, 36(3): 49-55, 122.

[3] CARLOS C, MARCELO M, BARBARA P. Information credibility on twitter [C]//Proceedings of the 20th international conference on World Wide Web. New York: ACM, 2011: 675-684.

[4] YANG F, YU X, LIU Y, et al. Automatic detection of rumor on Sina Weibo [C]//Proceedings of the ACM SIGKDD Workshop on mining data semantics. New York: ACM, 2012: 1-7.

[5] He Gang, Lü Xueqiang, Li Zhuo, et al. Research on Weibo Rumor Identification [J]. Library and Information Service, 2013, 57(23): 114-120.

[6] Xia Song, Lin Rongrong, Liu Kan. Research on Construction of Online Rumor Sensitive Word Lexicon: A Case Study of Sina Weibo Rumors [J]. Knowledge Management Forum, 2019, 4(5): 267-275.

[7] Li Gang, Wang Yuda. Research on Novel Coupled Social Network Rumor Propagation Model Based on Audience Profiling [J]. Journal of Modern Information, 2020, 40(1): 123-133, 143.

[8] KWON S, CHA M, JUNG K, et al. Prominent features of rumor propagation in online social media [C]//2013 IEEE 13th international conference on data mining. Dallas: IEEE, 2013: 1103-1108.

[9] MA J, GAO W, WEI Z, et al. Detect rumors using time series of social context information on microblogging websites [C]//ACM international conference on information and knowledge management. New York: ACM, 2015: 1751-1754.

[10] Wang Zhihong, Guo Yi. Research on Automatic Detection of Weibo Rumor Events [J]. Journal of Chinese Information Processing, 2019, 33(6): 132-140.

[11] KOTTETI M, DONG X, QIAN L. Multiple time-series data analysis for rumor detection on social media [C]//2018 IEEE international conference on big data. Seattle: IEEE, 2018: 4413-4419.

[12] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks [C]//Proceedings of the 25th international joint conference on artificial intelligence. New York: AAAI Press, 2016: 3818-3824.

[13] LI L, CAI G, CHEN N. A rumor events detection method based on deep bidirectional GRU neural network [C]//IEEE international conference on image. Piscataway: IEEE, 2018: 755-759.

[14] Wang Xingyu. Rumor Identification System Based on Long Short-Term Memory Network and Deep Features [D]. Baoding: Hebei University, 2020.

[15] Liu Kan, Du Haochen. Research on Twitter Rumor Detection Based on Deep Transfer Network [J]. Data Analysis and Knowledge Discovery, 2019, 3(10): 47-55.

[16] SIVA G, DEEPAK P, CHENG L, et al. Unsupervised fake news detection: a graph-based approach [C]//Proceedings of the 31st ACM conference on hypertext and social media. Florida: ACM, 2020: 75-83.

[17] MARRA F, GRAGNANIELLO D, COZZOLINO D, et al. Detection of GAN-generated fake images over social networks [C]//Proceedings of 2018 IEEE

conference on multimedia information processing and retrieval. Piscataway: IEEE, 2018: 384-389.

[18] QIAN F, GONG C, SHARMA K, et al. Neural user response generator: fake news detection with collective user intelligence [C]//Proceedings of the 27th international joint conference on artificial intelligence. Stockholm: IJCAI, 2018: 3834-3840.

[19] Fan Rong. Research on Rumor Propagation Characteristics and Behavior Prediction in Social Media During Public Emergencies [D]. Dalian: Dongbei University of Finance and Economics, 2019.

[20] Zeng Ziming, Wang Jing. Research on Weibo Rumor Identification Based on LDA and Random Forest: A Case Study of 2016 Haze Rumors [J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(1): 89-96.

[21] Wang Lin, Wang Ke, Wu Jiang. Public Opinion Dissemination and Evolution of Public Health Emergencies on Social Media: A Case Study of the 2018 Vaccine Incident [J]. Data Analysis and Knowledge Discovery, 2019, 3(4): 42-52.

[22] Li Lihua, Han Sining. Research on Network Public Opinion Propagation Mechanism and Prevention of Terrorist Attacks: An Empirical Analysis of Typical UK Cases [J]. Journal of Intelligence, 2019, 38(11): 102-111, 54.

[23] Cao Juan, Zhang Yongdong, Li Jintao, et al. A Density-Based Adaptive Optimal LDA Model Selection Method [J]. Chinese Journal of Computers, 2008(10): 1780-1787.

Author Contributions

Shi Kaiwen: Data acquisition, experimental programming, and initial manuscript drafting.
Liu Kan: Problem formulation, research design, and manuscript finalization.

English Abstract

Weibo Rumor Identification in Public Health Emergencies

Shi Kaiwen, Liu Kan
School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073

Abstract: [Purpose/significance] In public health emergencies such as the COVID-19 epidemic, a large number of statements about the epidemic have been quickly generated on social media on the Internet, including many rumors that endanger public mental health and affect the implementation of national policies. Detecting these remarks and identifying the rumors can enable the people to respond to public health emergencies correctly, and play a positive role in maintaining social stability and network governance. [Method/process] Firstly, the confirmed rumors during the epidemic were collected for in-depth analysis, and the main features of the rumor text were extracted, including context features, topic category features, sentiment level features, keyword features, etc.; then aiming at the problem that the text feature expression in the text classification model was relatively single, different models were used to vectorize the extracted rumor text features, and various text features were strengthened and fused. Among them, the word vector weight calculated by TF-IDF can strengthen the keyword feature information of word granularity while capturing the context features. Finally, the BiLSTM+DNN model was used to classify the fused feature vectors. [Result/conclusion] The experimental results show that features such as topic category and sentiment level all contribute to rumor recognition, especially the fusion of the strengthened word vector and other features significantly improves the recognition accuracy, and the indicators such as recall rate and F1 value all reach more than 90%, and the effect exceeds other rumor recognition models, indicating that the method constructed in this paper can respond well to the task of rumor recognition in the context of public health emergencies.

Keywords: public health emergencies; rumor recognition; Weibo; multi-feature fusion