RESEARCH

Open Access



Harmfulness metrics in digital twins of social network rumors detection in cloud computing environment

Hao Li¹, Wu Yang^{1*}, Wei Wang¹ and Huanran Wang¹

Abstract

Social network rumor harm metric is a task to score the harm caused by a rumor by analyzing the spreading range of the rumor, the users affected, the repercussions caused, etc., and then the harm caused by the rumor. Rumor hazard metric models can help rumor detection digital twins to understand and analyze user behaviors and assist social network network managers to make more informed decisions. However, there is a lack of models that can guantify the harm of rumors and automated harm metric models in rumor detection digital twins. To address this issue, this paper proposes an innovative social network rumor harm metric based on rumor propagation knowledge and a large language model (LLM), RSK-T5. The method first completes the joint task of rumor comment stance detection and sentiment analysis to capture critical features of rumor propagation. Then, this knowledge is used in the pretraining process of LLM to improve the model's understanding of rumor propagation patterns. Finally, the fine-tuning phase focuses on the hazard metrics task to improve the generalization energy. We compare with some existing variants of rumor detection methods, and experimental results demonstrate that RSK-T5 achieves the lowest MSE scores on three well-known rumor detection datasets. The ablative learning work demonstrates the effectiveness of RSK-T5's knowledge of two rumor spreads.

Keywords Rumor hazard metric, Digital twins, Large language model, Rumor propagation, Rumor stance detection, Rumor sentiment analysis, Social network

Introduction

The popularity of the Internet and social media has made it easier to disseminate information. However, it has also led to the spread of rumors and false information. These rumors may have severe negative impacts on individuals and society. Therefore, there is an urgent need to strengthen research on social network rumors to implement effective intervention strategies [1-3]. A promising solution is a digital twin system for rumor detection in a cloud computing environment [4, 5]. It can be used

*Correspondence:

yangwu@hrbeu.edu.cn

College of Computer Science and Technology, Harbin Engineering

University, Harbin 150001, Heilongjiang, China

to process large-scale rumor data in real-time, maintain model training and updating, ensure data security and privacy, and guarantee the ease of use and maintainability of rumor detection models. The left half of Fig. 1 lists the flow of the rumor detection digital twin system for social networks in cloud computing, which mainly contains the data, model, and application layers.

At the data and model level, current research involving rumors has mainly focused on the field of rumor detection to identify and filter rumors. Rumor detection tasks are mainly concerned with binary/trivial classification of a given information or message, i.e., determining whether it is a rumor or not [1]. However, this type of task does not give finer-grained categories of rumors with their modes of propagation and the harm that rumors can cause. The study of rumor harm



© The Author(s) 2024. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

Wu Yang



Fig. 1 The process of a digital twin system for social network rumor detection in the cloud with the problem addressed in this paper, using a set of real rumor propagation data (from the RumourEval dataset) as an example. The black arrows represent the direction of rumor propagation, while the red arrows indicate the direction of user stance and sentiment

and propagation patterns can further quantify rumors' impact, thus providing a basis for targeted interventions by social network managers. We define the social network rumor hazard metrics task as follows: given a rumor in a social network, its potential harm to society and individuals is predicted by analyzing its propagation pattern.

Research on the harmfulness of rumors spans several disciplines, such as psychology, law, and health care [6-9]. Nonetheless, evaluating the harm of rumors in these studies still focuses on determining the scope of a rumor's impact and the degree of negative emotions it provokes on social networks. However, there still needs to be a sufficient number of dimensions that define the impact of rumors. As shown in the right half of the case in Fig. 1, after completing the data acquisition, User1 questioned the rumor information. User3 supported User1's opinion in the same tone, proving they did not believe the rumor. However, the sentiment polarity of their comments and rumors was defined as negative, indicating that their sentiment were affected by the rumors. Rumor detection based on a single comment stance or sentiment information may lead to an incomplete grasp of rumor characteristics.

Meanwhile, in terms of public opinion analysis of the rumor detection digital twin system, automated rumor harm metrics studies still need to be improved to meet the real-time demand for harm prediction. Most studies wait until the rumor is completely finished spreading before the harm of that rumor can be analyzed. The detection system inevitably contains many a posteriori features. Therefore, we need to start from the information available in the short term, use the predicted a posteriori features as a jumping-off point, obtain the hazard annotation by calculating the historical data, and then train the end-to-end automatic prediction model.

This paper focuses on two main perspectives to overcome the above data and modeling layer challenges. First, the rumor hazard metric task relies more on rumor propagation information than the rumor detection task. For example, many rumor detection tasks can achieve good results using only multimodal information such as rumor text [10, 11] or images [12, 13]. The measure of the harm of a rumor not only needs to consider the degree of falsity and malice of the rumor itself but also to investigate its widespread and the public opinion reaction of the affected users in multiple dimensions. Therefore, fine-grained knowledge of rumor dissemination is essential in this study. By comprehensively analyzing the stance and sentiment of users' comments affected by rumors [14], we can better understand users' attitudes toward rumors and the dynamics of rumor spreading in social networks [15]. This interactive knowledge of the extent of rumor propagation, stance, and sentiment helps us to construct richer and more effective feature representations that can improve the performance of harm metrics. Second, in this study, we perform hazard metrics based on the Large Language Model (LLM) because the LLM has robust representation learning and transfer learning capabilities. Through pre-training and fine-tuning, LLM can effectively capture the complex patterns and features in rumor propagation, thus improving the accuracy of the hazard metric.

This paper proposes a social network rumor hazard metric RSK-T5 (Rumor Stance and Sentiment Knowledge based T5) based on Rumor Spreading Knowledge and Large Language Model (LLM). Precisely, we first extract the combined information of stance and sentiment of rumor comments from a large amount of social network data to capture critical features of the rumorspreading process. These features help the model to understand the real intention behind the comments more accurately. These two rumor propagation features were represented by two classification tasks: rumor comment stance detection and sentiment analysis. We then use this rumor propagation knowledge in the pre-training process of a large-scale language model to enable the model to understand better and analyze rumor propagation patterns. The fine-tuning phase then focuses on the hazard metrics task to improve the model's generalization ability and relevance. By using RSK-T5, we can more effectively measure the harm of rumors in social networks and help digital twins understand the impact of rumors in real-time, thus providing decision-makers with powerful data support and action guidelines.

The contributions of this paper can be summarized as follows:

- (1) We propose a rumor harmfulness metric based on a large language model, which can meet the real-time demand for rumor harmfulness metrics in rumor detection digital twin systems. To our knowledge, this is the first automated method for detecting rumor hazard.
- (2) We use a pre-training mechanism based on rumor comment stance and sentiment information, which can comprehensively consider the background knowledge of stance classification and sentiment analysis tasks, learn finer-grained knowledge of rumor propagation, and support downstream rumor hazard metrics tasks.
- (3) Experimental results on three real-world datasets show that the method in this paper outperforms several existing variants of rumor detection methods. Ablation learning is proposed to prove the effectiveness of rumor propagation knowledge. We also analyze which granularity of rumor propagation knowledge would be helpful for the hazard metrics task.

Related work

There has been a great deal of research on rumor hazards. This section presents three main areas relevant to this paper: quantitative research on rumor hazards, rumor comment stance detection and sentiment analysis, and text-to-text large language modeling.

Rumor hazard metrics

In order to systematically examine the hazards of misinformation, Tran et al. [16] proposed a Conceptual Model for Mitigation of Misinformation Harms during Crises using activity theory. This framework enables interaction between humans and machines and their respective loops. Their work presents for the first time the task of mitigating the harms of misinformation.

Greenspan et al. [8]. investigated the consequences of misinformation dissemination during the Covid-19 epidemic. The authors claim that it affects people's beliefs about the effects of the disease, preventive behaviors, and even their memories of their past experiences. Kostkova et al. [9] present a project called VAC Medi + Board, which aims to visualize the spread of rumors through social networks and assess key people's impact. Finally, it measures the level of hesitation about vaccines. Sandor [7] explores the impact of rumors on the intervention logic and practice of the United Nations Multilateral Stabilization Mission in Mali (MINUSMA), which argues that rumors show an epistemological power that can have an intrinsic effect among the different actors of the conflict in Mali.

Wang et al. [17] investigated the damage of social network rumors on trust and provided insights into trust building. Castillo et al. [18] determined the credibility of rumors by automatically categorizing rumors as credible and untrustworthy. This is the beginning of rumor detection research, which, on the other hand, supports the task of rumor harm detection.

Rumor hazard research spans across emerging technologies such as economic management [19], healthcare [9], international relations [7], physics [20], and human–computer interaction [17] [8, 19]. However, there needs to be more automated rumor hazard metrics research to ensure real-time assessment and analysis of rumor hazards from an applied systems perspective. Moreover, there have been many successful applications of digital twin technology based on social network analysis [21, 22]. Harmfulness metrics based on the knowledge of rumor spreading in social networks can also provide a real-time monitoring and information assessment mechanism for the digital twin system.

Rumor stance detection and sentiment analysis Rumor stance detection

The task of rumor stance detection is to determine the user's stance towards a rumor in rumor propagation, which can help to reveal the rumor propagation mechanism and assess the impact of the rumor. Zubiaga et al. [23] investigated the performance of sequential classifiers using social media conversational cues on the task of classifying rumor stance and tested four sequential

classifiers (Hawkes process, Linear Chain Conditional Random Fields, Tree Structured Conditional Random fields, and long and short-term memory networks) on eight datasets related to breaking news. The results show that sequential classifiers using local features outperform non-sequential classifiers and that LSTM with a simplified feature set performs best. [24] The authors propose a feature-rich stacked LSTM model with new F1 metrics by comparing features and architectures.

Regarding the interaction between rumor detection and stance detection, Ma et al. [25] proposed a deep neural network-based multi-task learning framework to jointly process rumor detection and stance classification tasks to improve both performances. [26] delved into feature engineering for stance classification in social media rumor detection and screened out 18 salient features involving text content, user profiles, and dissemination status. A traditional logistic regression classifier using these 18 features achieved state-of-the-art performance on the RumorEval dataset.

Recently, it was found by researchers that sentiment features can provide support for rumor stances. [27] proposed a new method to predict users' stance against rumors on Twitter by using conversation- and sentimentbased features. Hardalov et al. [28], on the other hand, conducted a cross-linguistic stance detection study using the pattern exploitation training (PET) method. An F1 improvement of more than 6% relative to the baseline in a few-shot learning environment was achieved by introducing a new tag encoder and a sentiment-based stance data generation strategy.

Rumor sentiment analysis

There has been early research on sentiment in social media. Godbole et al. [29] assigned positive or negative opinion scores to each entity in a text corpus through a sentiment recognition and sentiment aggregation scoring phase. The importance of scoring techniques was evaluated on a large corpus of news and blogs. Pang et al. [30] investigated techniques and approaches that directly support opinion-oriented information retrieval systems, focusing on the new challenges posed by sentiment-aware applications. The authors cover broader issues such as summarization of evaluative texts, privacy, manipulation, and economic impact.

In recent years, research on neural network-based sentiment analysis of neural networks has gradually become a research focus. Wang et al. [31] proposed a novel twotier cascaded gated recurrent unit (CGRU) model based on a sentiment lexicon and a two-step dynamic timeseries algorithm for detecting rumor events in online social networks. [32] proposed a novel three-stage process to detect the stance of tweet authors, including preprocessing, feature generation, and classification. The innovative feature selection was achieved using two ranked lists based on word frequency-inverse document frequency (tf-idf) scores and sentiment information. In the work on using rumor sentiment features to aid rumor detection, [33] proposed a multimodal dual sentiment feature for rumor detection, including published visual sentiment, published text sentiment, and social sentiment. Experimental results proved that image sentiment improved rumor detection efficiency. The proposed multimodal dual sentiment feature can be used as a model plug-in to integrate with various rumor detection models to improve performance seamlessly. Hossain et al. [34] proposed a framework for automatic sentiment recognition systems based on IoT, edge computing, and cloud computing to reduce the computational burden on the client and ensure the user's privacy. Ghorban et al. [35] addressed the issue of the high cost of sentiment analysis when dealing with large-scale data in cloud computing scenarios.

Summary

Although the research on stance detection and sentiment analysis of rumor comments provides excellent help to the rumor detection problem, there is no current research on modeling rumor conversations that integrates consideration of comment stance and sentiment. As shown in the case of Fig. 1, a single coarse-grained rumor comment feature can lead to limited model functionality.

Text-to-text large language modeling

The pre-training mechanism provided by the Big Language Model uses less training costs and allows for rapid deployment and immediate use [35, 36]. Large language models have become the focus of research on NLP tasks in recent years by demonstrating strong text comprehension and generation capabilities when dealing with NLP problems [37]. Raffel et al. [38] explored the full spectrum of NLP transfer learning techniques by introducing a unified framework, T5, for converting all text-based language problems into text-to-text format. The authors' study compares pre-training goals, architectures, unlabeled datasets, migration methods, and other factors for dozens of language understanding tasks. This work gives the entire field of NLP pre-training models a generalized framework that translates all tasks into one form. Ni et al. [39] recently proposed the SENTENCE T5 model, wanting to explore the capabilities of T5 in text representation. The authors found that the encoder output of hidden doing average pooling can have good results. Moreover, T5 can not only be used for NLU tasks as a whole, but the T5 encoder can also outperform BERT



Fig. 2 Model structure

and achieve good results in downstream tasks.Flan-T5 [40] is the latest work from Google, giving the language model extreme generalization performance by fine-tuning it on mega-scale tasks. Once again, researchers have advanced the performance level of instruction tuning.

The work in this paper can be viewed as an extension of the T5 model on the specific task of rumor hazard metrics. Based on the flexibility of the T5 model, we added two kinds of rumor-spreading knowledge: stance and sentiment. The generalization ability and performance of the model are improved.

Approach

In this section, we introduce our model for rumor hazard metrics. As shown in Fig. 2, the rumor hazard metric problem consists of two subtasks: 1) a pre-training task based on the stance and sentiment knowledge of rumor comments, and 2) a rumor hazard metric task based on a fine-tuned regression model. We present the details of each component in the subsequent subsections.

Problem statement

Let the rumor dataset be $\mathbb{D} = \{R_1, R_2, \dots, R_{|R|}\}, R_i$ represents the i-th rumor and the comments surrounding it, $Y^h = \{Y_1^h, Y_2^h, \dots, Y_{|R|}^h\}$ and $Y_i^h \in \mathbb{R}$ denotes the harm score of the rumor. Each $R_i = \{r_i, c_1, c_2, \dots, c_{|R_i|}\}, c_i$ represents a comment. For a comment text c, it has the label $Y_c = \{Y_c^{St}, Y_c^{Sen}\}$. Where $Y_c^{St} \in Y^{St} = \{Y_1^{St}, Y_2^{St}, \dots, Y_{|R_i|}^{St}\}$ represents the stance label of c, which takes the values: support, deny, question, and simple comment (SDQC). $Y_c^{Sen} \in Y^{Sen} = \{Y_1^{Sen}, Y_2^{Sen}, \dots, Y_{|R_i|}^{Sen}\}$ represents the sentiment label of c, taking the values: positive, negative, neutral. We hypothesize that when a model can accurately distinguish between users' stance and sentiment towards a rumor in a rumor comment and can accurately classify the two, it will have a more vital ability to identify the harmfulness of this rumor. This is because, at this point, the model can discover more fine-grained features of rumor propagation, which serves the same purpose as traditional manual feature extraction for rumors.

As presented in Sect. 1, our goal is to learn a model that has knowledge of the finer granularity of rumor propagation. This includes stance knowledge and sentiment knowledge of rumor comments.

Let \mathcal{D}_{harm} be the original rumor training set containing the rumor texts and their harm scores. Let \mathcal{D}_{knowl} be the rumor comment training set, and \mathcal{D}_{harm} and \mathcal{D}_{knowl} be split from the dataset \mathbb{D} . Our goal is to learn a model \mathcal{M} makes:

 \mathcal{D}_{pre} denotes a large-scale pre-trained dataset, and θ_1 is a pre-trained model on \mathcal{D}_{pre} . θ_2 is a model trained on $\mathcal{D}_{\text{knowl}}$, and θ_1 determines its initialization parameters.

$$\arg \max_{\mathcal{M}} \log p(\mathcal{M}\mathcal{D}_{\text{harm}}, \mathcal{D}_{\text{knowl}}, \mathcal{D}_{\text{pre}}) = \mathcal{M}$$

$$\arg \max_{\mathcal{M}} \left[\log p(\mathcal{M}\mathcal{D}_{\text{harm}}, \theta_2) + \log p(\theta_2 \mathcal{D}_{\text{knowl}}, \theta_1) + \log p(\theta_1 \mathcal{D}_{\text{pre}}) \right]$$

$$\mathcal{M}$$
(1)

Model structure

T5 Model

T5 (Text-to-Text Transfer Transformer) is a pre-trained language model based on the Transformer architecture. The core idea of the T5 model is to unify all natural language processing tasks into a single text-to-text transformation task. This approach allows the model to be pre-trained and finetuned on the rumor-hazard metric task, thus enabling end-to-end transfer learning.

The T5 model consists of an Encoder and a Decoder consisting of multiple Transformer layers, respectively. The Encoder is responsible for encoding the input text sequence into successive hidden states, and the Decoder generates the target text sequence based on the output of the Encoder. Each Transformer layer consists of a multihead self-attentive sublayer and a feed-forward neural network sublayer. These sublayers both contain residual connectivity and layer normalization.

$$\mathbf{p}_{k} = \mathrm{T5}(s, t_{1:k-1}) = Softmax(Dense(Dec(Enc(s), t_{1:k-1})))$$
(2)

In this paper, we are concerned with the semantic information of the input text rather than generating a new text sequence. Therefore, it is sufficient to use the encoder of the T5 model to extract the semantic information of the text. Specifically, T5 encapsulates each Transformer layer as a block, and then superimposes multiple blocks to form the encoder.

Pre-training based on rumor comment knowledge

The purpose of model θ_2 is to make full use of the knowledge of the T5 model and learn features related to the two classification tasks based on it. Namely, solving arg max_{θ_2} log $p(\theta_2 D_{\text{knowl}}, \theta_1)$. We first obtain a model for the rumor comment stance detection and sentiment analysis tasks by finetuning based on the T5 pre-training model. For a rumor r_i , the category of its comment c_{ij} is:

$$\hat{\mathcal{Y}}_{ij} = f(Dropout(Enc(c_{ij}, r_i)))$$
(3)

For the comment stance detection task and the comment sentiment analysis task, we define two linear classification functions $f_1 : \mathbb{R}^D \to \mathbb{R}^{K_1}$, $f_2 : \mathbb{R}^D \to \mathbb{R}^{K_2}$, which are mappings of comment embedding vectors to classification category scores. Where $K_1 = 4$, represents the four SDQC stance categories; $K_2 = 3$, represents the three PNC sentiment categories. Where each of $f(x_i, W, b) = Wx_i + b$.

Rumor hazard metric task finetuning

Rumor hazard metric task finetuning

Next, we use this classification model as a pre-trained model and finetune it again for the regression task. That is, solving the arg max $\log p(\mathcal{MD}_{harm}, \theta_2)$.

Based on model θ_2 , the harm score of a rumor r_i can be expressed as:

$$\hat{\mathcal{Y}}_{i} = f(Dropout(Enc(Embed(r_{i}, \lambda_{2}))))$$
(4)

 λ_2 is a parameter learned by model θ_2 on the data $\mathcal{D}_{\text{knowl}}$. λ_2 represents rumor r_i 's comment stance and knowledge information. The structure of the finetuned model is the same as Eq. (3), as a linear scoring function, $f_3 : \mathbb{R}^D \to \mathbb{R}^{K_3}$, and $K_3 = 1$. In this way, our regression model can inherit rumor propagation knowledge from the classification model, improving the regression task's performance.

Model details

The input comment text *c* or rumor text *r* is first tagged as word fragments: $c = \{c_1, c_2, ..., c_m\}$, $r = \{r_1, r_2, ..., r_n\}$. Then, we input the token sequence into the word embedding layer *Embed*() to obtain the input embedding X_c or X_r . The encoder is based on the T5 structure, and its Transformer block can be represented as a combination of multi-head attention and feed-forward neural network. Our encoder uses a multi-head self-attention mechanism to capture information for each Transformer block's upper input layer \mathbf{H}_{k-1} . For comment information, *Att* is used to capture semantic features specific to rumor stance and sentiment; for rumor information, it is used to capture stance and sentiment knowledge related to rumor harm.

$$\mathbf{H}_{k}^{\prime} = Att\left(Norm\left(\mathbf{H}_{k-1}\right)\right) + \mathbf{H}_{k-1} \tag{5}$$

where *Norm*() denotes layer normalization. Then, \mathbf{H}'_k is input *FNN*() to get the feature representation of the current Transformer block:

$$\mathbf{H}_{k} = FNN(Norm(\mathbf{H}_{k}')) + \mathbf{H}_{k}'$$
(6)

Finally, the output of the encoder passes through the Dropout layer and the Linear layer f to obtain the knowledge classification result or rumor hazard score \hat{y} .

Training

To investigate the effect of rumor comment data, we explored two stages of training: first, training on the rumor comment stance detection task and the sentiment analysis task; then finetuning the model using manually annotated labels of the original rumor hazard scores.

Specifically, for the input *x* output *y*, The two pre-training tasks loss:

$$L(\theta_2)^{St} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^4 y_{ij}^{St} \log p_{ij} + \frac{\lambda_1}{2} ||\theta_2|^2$$

$$L(\theta_2)^{Sen} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^3 y_{ij}^{Sen} \log p_{ij} + \frac{\lambda_2}{2} ||\theta_2|^2$$
(7)

where p_{ij} is the predicted probability that the observed sample *i* belongs to the category *j*, $y_{ij}^{St} \cdot y_{ij}^{Sen}$ is the sign function, $\lambda_1 \cdot \lambda_2$ is the L2 regularization parameter, and θ_2 is the set of parameters of the pre-trained model. The total loss optimized by the model for learning rumor comment knowledge is $L(\theta_2) = L(\theta_2)^{Sen} + L(\theta_2)^{St}$.

In the finetune phase of the task for the rumor hazard metric, the loss of the task can be expressed as:

$$L(\mathcal{M}) = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$
(8)

Automatic harmfulness labeling

Due to the unavoidable subjectivity and lack of scalability of manual hazard labels, we first pre-label the data using an automated rule-based annotation method. Inspired by previous work [41-43], our annotation method focuses on three metrics: sentiment, identity, and the degree of rumor harm and organization.

We propose a rumor sentimentality R_c to reflect the intensity of negative emotions triggered by rumors. Rumor approval R_r is proposed to reflect the support level from other users for this rumor. The formula is shown in Eq. (9), where C_n and C_p represents the negative and positive ratings of the rumor and, $S_s \\ S_d \\ S_d \\ represents the number of comments supporting, opposing, and questioning the rumor.$

$$R_{c} = \frac{\sum_{U} w_{u} \times C_{n}}{\sum_{U} w_{u} \times C_{n} + \sum_{U} w_{u} \times C_{p}}$$

$$R_{r} = \frac{\sum_{U} w_{s} \times S_{s}}{\sum_{U} w_{s} \times S_{s} + \sum_{U} w_{d} \times S_{d} + \sum_{U} w_{q} \times S_{q}}$$
(9)

The next indicator R_o to be considered is the degree of organization of the rumor. The above study first calculates the distribution of the number of people obtained from the participating users sorted by the number of retweets. Then it calculates the skewness to see whether the rumor is widely spread.

$$R_o = \frac{\text{Skew}(UF_R) - \text{Skew}(\overline{UF_{NR}})}{\text{Skew}(UF_R)}$$
(10)

Experiments

Data collection

Our experiments were conducted on two datasets containing the original rumor text and several comments against it. Rumor tags and comment stances are tagged into the data. The details of the datasets are in Table 1. One of them, PHEME, is a classical rumor and stance detection dataset containing 297 original rumors with 4589 comments surrounding it. It should be noted that the PHEME dataset's original rumor comments are labeled as agreed, disagreed, comment, and appealfor-more-information. For ease of representation, we denote it uniformly with the RumourEval dataset as sdqc. rumourEval is a dataset from the dataset from Semeval-2019 on the text classification task of rumors. The classification task contains rumor classification and rumor stance classification. The RumourEval dataset contains 446 rumors and 7995 comments from Twitter and Reddit. Since the Reddit dataset is small, we combine the two in our experiments. The FNC dataset [25] is released by the 2017 Fake News Challenge on news articles and its comments. Among them, comment tags include: agrees, disagrees, discussions and unrelated.

After pre-labeling using the method in Sect. 3.4 for rumor harm labels, we collected corrected harm scores from 10 volunteers. We took the average of the ten scores as the harm score of this comment. We use the vaderSentiment method [44] for the sentiment classification label to obtain the results of one of the three classifications.

Since the FNC data set is relatively large, we use a mixture of manual and automatic annotation to add labels to it. The a priori rules are as follows:

Dataset	Comment	t Stance			Commen	t Sentiment		Source l	Rumor		
PHEME	Support	Deny	Query	Comment	Positive	Negative	Neural	Rumor		Non-rumor	
	683	371	398	3137	.239	297	4053	297		0	
RumourEval	Support	Deny	Query	Comment	Positive	Negative	Neural	True	False		Unverified
	821	546	553	6075	.409	455	7131	185	138		123
FNC	agree	disagree	discuss	unrelated	Positive	Negative	Neural	Fake nev	VS		
	5581	1537	13,373	54,894	133	2744	72,508	2559			

 Table 1
 Statistics of the datasets

Prior knowledge of data annotation

1 When the comment stance is comment, support, or the comment sentiment has the same polarity as the original rumor, it can be understood as supporting the spread of the rumor. The more such tags, the greater the harm of the rumor

2 When the comment stance is deny and query, or the comment sentiment is of a different polarity than the original rumor, it means there is a rumor refuting information. The more such tags, the less harmful the rumor is

3 The depth and breadth of the rumor propagation tree and the total number of replies can reflect the degree of spread of the rumor. The degree of spread of the rumor is positively related to the degree of harm of the rumor

4 The greater the number of followers and friends of a user who participates in the rumor, the more significant his contribution to the spread of the rumor. Its commentary stance and sentiment are more important than those of ordinary users

5 The original post with the rumor label true rumor or non-rumor is, by default, less harmful than rumor

Experimental setup

The version of T5 we use is the Flan-T5-Large model (780 M parameters). Flan-T5 is a T5 model with a more robust generalization performance obtained by a multi-tasking fine-tuning scheme (Flan). In order to test the performance of the model, we compared the method of this paper with several baselines. Since no current work on rumor hazard metrics exists, our comparison method contains some deep-learning prototype algorithms or variations of current rumor detection methods. Baseline specifically includes:

- (1) a Bert-LSTM-based approach: This is a variant of the [45] approach. We use the BERT model for pre-training on rumor comment data. The Bert model is frozen in the rumor hazard metric stage so that it only generates comment expressions, and then the LSTM model is used to obtain the hazard score.
- (2) Bert-based approach: The Bert model has the same encoder-only structure as the text-based approach. However, the T5 model has no CLS token at the beginning of each sentence, and the pre-training phase is a "generative pre-training task."
- (3) BERT-TextCNN2:. A variant of the rumor detection method [46]. The model structure is similar to BERT-LSTM, and we use a convolutional neural network model as a regression task for the rumor hazard metric.
- (4) ablation experimental model: We also constructed several model variants for ablation analysis, including RSK-T5-Base

 RSK-T5-Base
 RSK-T5-sentiment
 RSK-T5-stance.

In the pre-training phase of extracting rumor comment knowledge, we used accuracy, precision, recall, and F1 scores to evaluate the classification results of rumor comment stance detection and sentiment analysis. These classification results are used to see if the model has enough information about the impact of rumors. As with the loss function, we use the mean square error (MSE) to measure the model's performance in predicting rumor harm.

Table 2 Experimental results on the PHEME dataset

	MSE	Stance-F1	Stance- Accuracy	Stance- Precision	Stance-Recall	Sentiment-F1	Sentiment- Accuracy	Sentiment- Precision	Sentiment- Recall
BERT-LSTM	0.117	0.338	0.679	0.378	0.438	0.66	0.894	0.626	0.707
BERT-TextCNN	0.129	0.466	0.684	0.447	0.514	0.693	0.91	0.646	0.765
BERT _{ft}	0.113	0.471	0.653	0.451	0.516	0.789	0.935	0.737	0.865
RSK-T5	0.105	0.481	0.656	0.465	0.503	0.81	0.939	0.775	0.853

Table 3 Experimental results on the RumourEval dataset

	MSE	Stance-F1	Stance- Accuracy	Stance- Precision	Stance-Recall	Sentiment-F1	Sentiment- Accuracy	Sentiment- Precision	Sentiment- Recall
BERT-LSTM	0.121	0.48	0.777	0.453	0.555	0.67	0.902	0.654	0.692
BERT-TextCNN	0.154	0.456	0.791	0.423	0.624	0.636	0.911	0.56	0.824
BERT _{ft}	0.101	0.518	0.765	0.51	0.534	0.757	0.93	0.728	0.934
RSK-T5	0.069	0.56	0.766	0.577	0.549	0.75	0.928	0.717	0.792

	MSE	Stance-F1	Stance- Accuracy	Stance- Precision	Stance-Recall	Sentiment-F1	Sentiment- Accuracy	Sentiment- Precision	Sentiment- Recall
BERT-LSTM	0.069	0.762	0.944	0.749	0.802	0.669	0.985	0.661	0.917
BERT-TextCNN	0.104	0.866	0.976	0.854	0.884	0.999	0.999	1.0	0.999
BERT _{ft}	0.078	0.209	0.721	0.25	0.18	0.327	0.963	0.333	0.321
RSK-T5	0.061	0.987	0.996	0.986	0.987	1.0	1.0	1.0	1.0

Table 4 Experimental results on the FNC dataset

Results of the experiments

Main experiment

Tables 2, 3 and 4 show the results of our experiments on the two datasets. Among them, the F1, accuracy, precision, and recall of rumor comment stance detection and sentiment analysis are obtained in the pre-training phase. And MSE is obtained by fine-tuning in the rumor hazard metric task. The experiments and analysis can be summarized as follows:

- The RSK-T5 model achieved the best MSE scores in both datasets, which proves the effectiveness of the method in this paper.
- (2) The BERT_{ft} model based on pre-training and finetuning of rumor propagation knowledge also achieved a lower MSE score, which was only 7.62%, 46.3% and 27.9% higher than the RSK-T5 model on the three datasets. This indicates that the fine-tuned model has a better adaptation to the rumor hazard metric task.
- (3) The models BERT-LSTM and BERT-TextCNN constructed by freezing BERT after pre-training demonstrated higher MSE scores, which are on average 33.3% and 72.2% higher than the RSK-T5 model in the three datasets. This indicates that the regression models based on LSTM and TextCNN cannot make better use of rumor propagation knowledge. The experimental results also prove the importance of fine-tuning.
- (4) Interestingly, our approach does not achieve the best performance on rumor comment stance classification or sentiment analysis alone. For example, the accuracy of BERT-TextCNN model for stance categorization on both datasets is higher than that of RSK-T5 method by 4.27% and 3.26%, respectively, and so on. These results indicate that the RSK-T5 model is a comprehensive consideration of the background knowledge of the two types of tasks, and ultimately aims to improve the performance of the rumor hazard metric.
- (5) In particular, the BERT_{ft} model achieved the best of all metrics on the rumor comment sentiment analysis task for the RumourEval dataset, yet the MSE

score was higher than the RSK-T5 model by 46.37%. This proves that a single comment sentiment analysis task cannot provide sufficient support for the rumor analysis task.

(6) An exception is that in the FNC dataset, the BERTft model achieved the worst rumor review stance detection and sentiment analysis performance. However, the MSE score of the BERTft model is still 25% lower than the BERT-TextCNN model. At the same time, although the RSK-T5 model achieved nearly 100% classification performance in the two tasks, its hazard measurement performance improvement was limited. These results suggest some gaps between the rumor hazard labels we constructed on the FNC dataset and their rich propagation characteristics.

Does knowledge of rumor comments help measure rumor harm?

In the previous subsection, we demonstrated the excellent performance of the RSK-T5 method in measuring rumor hazard, and we will investigate the role of rumor comment knowledge on rumor hazard. As shown in Fig. 3, we compare the performance of all the methods with and without rumor comment knowledge on both datasets. There is a substantial decrease in the MSE scores of the models when they use rumor comment knowledge. This again proves that rumor comment knowledge is additive to the performance of measuring rumor hazard.

A more specific example is on the RumourEval dataset, where BERT-LSTM has a minor boost using rumor comment knowledge. This may be because the sentence length of rumor comments in the RumourEval dataset is significantly higher than in the PHEME dataset. Compared to the BERT_{ft} and RSK-T5 models, BERT-LSTM is still not strong enough to deal with dependencies spanning a wide range of text sequences. Nevertheless, the performance of BERT-LSTM, BERT_{ft}, and RSK-T5 methods without using rumor comment knowledge is still better than that of BERT-TextCNN, which also indicates that



choosing a good regressor is crucial in measuring rumor hazards, in addition to making use of rumor comment knowledge.

The results of the BERT_{ft} model and the BERT-TextCNN model on the FNC dataset show that their performance decreases as they acquire more propagated knowledge. This may be because the FNC data set comprises extensive articles containing fake news rather than short message texts. Therefore, compared to the other two models, BERT-LSTM can be better utilized in the pre-training phase to learn more information about the structure of the language using the BERT model. On the other hand, this may mean some noise in the data in the pre-training stage of the FNC data set, which leads to a decrease in the performance of the BERT_{ft} model and the BERT-TextCNN model.

How much rumor comment knowledge is needed?

We will further explore what types of rumor comment knowledge critically impact the performance of the model's hazard measure. We classify different types of rumor comments in three dimensions: the depth of the comment, the breadth of the comment, and the text length of the comment itself. We then validate the model's demonstrated rumor hazard metric performance across all comments.

 Table 5
 RSK-T5
 model
 performance
 under
 different
 scenarios
 (MSE)
 Image: scenarios
 scenario
 scenario
 scenario

Dataset Condition	PHEME	RumourEval	FNC
 Depth≤3	0.168	0.131	
Depth > 3	0.078	0.105	—
Width≤12	0.132	0.171	0.08
Width > 12	0.099	0.122	0.039
Number of words ≤ 12	0.077	0.167	0.041
Number of words > 12	0.138	0.137	0.04

Our three rumor delineation criteria work as follows: rumor comment depth reflects the distance a message has traveled in the social network, i.e., the maximum number of retweets a message has gone through from the initial poster to the farthest receiver. Comment breadth reflects how far the message has spread in the social network, i.e., the maximum number of users the message can reach at a given propagation level. The number of words in a rumor comment, on the other hand, reflects the amount of information in the text.

Table 5 demonstrates the rumor hazard results obtained from the RSK-T5 model pre-trained using the full rumor propagation knowledge and fine-tuned at different rumor propagation scales. The RSK-T5 model significantly exhibits the best MSE scores for rumor comment depth greater than 3 and breadth greater than 12. A greater depth may imply that the rumor can spread and influence more remarkably. A greater breadth may indicate that the rumor has a higher spreading speed and coverage at a particular dissemination stage. Thus, this result suggests that the harm of a rumor is easier to measure when the rumor spreads more widely. However, when the number of comment words is greater than 12, the MSE score of our model increases by 79.2% on PHEME, decreases by 18% on RumourEval, and remained basically unchanged on the FNC dataset. This suggests that the rumor comment word count size does not clearly affect the rumor hazard metric task on the pre-trained model based on the full rumor propagation knowledge. This is because, in social networks, users are more accustomed to using short texts. Having longer text for comment information does not mean having more information.

In order to further explore the ability of the model to process a single type of data, we used the method of pre-training and fine-tuning in a single class of data, and the experimental results are shown in Tables 6, 7 and 8. The results and analysis can be summarized as follows:

	MSE	Stance-F1	Stance- Accuracy	Stance- Precision	Stance-Recall	Sentiment-F1	Sentiment- Accuracy	Sentiment- Precision	Sentiment- Recall
Depth≤3	0.179	0.479	0.651	0.484	0.544	0.666	0.89	0.602	0.815
Depth > 3	0.146	0.4	0.602	0.411	0.411	0.75	0.947	0.733	0.772
Width≤12	0.111	0.436	0.617	0.426	0.611	0.591	0.872	0.568	0.777
Width > 12	0.122	0.501	0.675	0.478	0.545	0.736	0.931	0.692	0.815
Words Num≤12	0.064	0.461	0.659	0.461	0.465	0.714	0.898	0.68	0.766
Words Num > 12	1.57	0.493	0.665	0.465	0.544	0.764	0.947	0.722	0.839

Table 6 Experimental results of the PHEME dataset under various rumor propagation scenarios

Table 7 Experimental results of the RumourEval dataset under various rumor propagation scenarios

	MSE	Stance-F1	Stance- Accuracy	Stance- Precision	Stance-Recall	Sentiment-F1	Sentiment- Accuracy	Sentiment- Precision	Sentiment- Recall
Depth≤3	0.158	0.61	0.753	0.583	0.647	0.802	0.923	0.771	0.843
Depth > 3	0.125	0.51	0.752	0.522	0.499	0.748	0.942	0.736	0.766
Width≤12	0.139	0.573	0.693	0.56	0.595	0.714	0.907	0.707	0.725
Width > 12	0.13	0.528	0.7621	0.517	0.546	0.754	0.931	0.713	0.809
Words Num≤12	0.097	0.554	0.703	0.55	0.562	0.783	0.922	0.801	0.766
Words Num > 12	0.102	0.458	0.767	0.455	0.465	0.703	0.952	0.666	0.756

Table 8 Experimental results of the FNC dataset under various rumor propagation scenarios

	MSE	Stance-F1	Stance- Accuracy	Stance- Precision	Stance-Recall	Sentiment-F1	Sentiment- Accuracy	Sentiment- Precision	Sentiment- Recall
Width≤12	0.069	0.936	0.959	0.925	0.948	0.63	0.991	0.99	0.625
Width > 12	0.069	0.988	0.997	0.988	0.989	1.0	1.0	1.0	1.0
Words Num≤12	0.069	0.986	0.996	0.986	0.986	0.999	0.999	1.0	0.999
Words Num > 12	0.036	0.979	0.993	0.979	0.979	1.0	1.0	1.0	1.0

- (1) The model's performance is similar in the case of different comment depths and breadths on the PHEME and RumourEval datasets. This indicates that when only one type of rumor exists in the social network, the model's ability to use its knowledge decreases. Disseminating knowledge of a broader range of rumors also fails to produce a significant performance gain for the hazard metrics task.
- (2) A remarkable phenomenon is that the model obtains the highest F1 score, accuracy, precision, and recall on the sentiment analysis task when processing data with a comment depth greater than 3 and a breadth greater than 12. This suggests that

users tended to exhibit more explicit sentiment polarity as comments around rumors unfolded. This phenomenon did not occur in the rumor comment stance detection task. User stance is often more difficult to confirm because as rumor participation in user conversations unfolds, the focus of user comments may shift from the rumor itself to some topic of greater interest to the user. Therefore, more information (e.g., where the user comments point) is needed to detect stance.

(3) In the rumor review data with different numbers of words, no clear pattern indicates the model's performance on the pre-trained stance detection, sen-



Fig. 4 Ablation study of model components on four datasets

timent analysis tasks, and fine-tuned rumor harmfulness measurement tasks. Specifically, when the RSK-T5 model processes the number of comment words in the PHEME data set is less than or equal to 12, the MSE performance is 145.3% higher than when the number of comment words is greater than 12. However, the stance detection and sentiment analysis performance needs to catch up. The performance on the FNC data set is exactly the opposite. The performance of the RSK-T5 model under different numbers of comment words on RumourEval is basically the same. This again validates our conclusions in Table 4 about the number of words in social network comments.

Ablation experiments

In order to investigate the effect of each module in the RSK-T5 model on the model performance, ablation analysis was proposed. As shown in Fig. 4, RSK-T5-Base, without any rumor communication knowledge background, achieved the highest MSE score. Secondly, the RSK-T5-sentiment method using only rumor comment sentiment knowledge demonstrated the best stance detection performance in both datasets. However, the MSE scores of the RSK-T5-sentiment are 7.6%, 44.9% and 24.6% higher than those of the RSK-T5 model, respectively. This phenomenon suggests that the sentiment knowledge of rumor comments alone cannot contribute much to the rumor hazard metric.

Secondly, the RSK-T5-stance method using only comment stance knowledge performs poorly in the PHEME and RumourEval datasets (MSE scores are 1.8% and 6% higher than RSK-T5-sentiment), and its rumor comment stance detection performance is lower than RSK-T5. This proves the complexity of rumor comment stance detection, which needs to synthesize more semantic information (not limited to sentiment knowledge) to achieve good results. Again, on the FNC dataset, all three models achieved very high stance detection and sentiment analysis accuracy. However, RSK-T5 can still utilize more complete knowledge of rumor propagation and achieve the lowest MSE score. Finally, the RSK-T5 model achieved the best MSE scores on both datasets, which proves the effectiveness of RSK-T5's rumor propagation knowledge.

Conclusions and future work

The popularity of the Internet and social media has led to a growing problem of rumor and disinformation dissemination. Rumor detection digital twins can effectively identify and assess rumors for social network platforms. However, current systems and research mainly focus on identifying rumors, and more work is needed for sufficiently large granularity to be done in assessing rumors in online social networks. To our knowledge, there has been little research on automated rumor hazard metrics. On the other hand, there has been much work on assisting rumor detection based on the stance and sentiment of user comments in rumor propagation. However, current methods are based on a single task of disseminating information at a coarsegrained level and do not take into account the interaction of stance and sentiment knowledge.

In this paper, we propose an automated rumor hazard metric RSK-T5 based on rumor propagation knowledge and a large language model from the data and model perspectives of a rumor detection digital twin system. The RSK-T5 model integrally considers the background knowledge of the stance categorization and sentiment analysis tasks, intending to improve the performance of the rumor hazard metric. The experimental results show that RSK-T5 outperforms existing rumor detection methods by an average of 26.7% on the three datasets regarding rumor hazard metrics. The experiments further validate that rumor comment knowledge enhances the performance of measuring rumor harm. We also explored the impact of different depths of propagation, breadth, and number of comment words on model performance and evaluated the effectiveness of rumor propagation knowledge at various granularities.

However, some things could be improved in our work. As a data-driven task, the granularity of propagation knowledge on which our harms are based still needs to be improved. Therefore, future work can collect more data after rumor propagation to see how rumors affect users' behaviors and decisions. In addition, in our future work, we will build the rumor hazard metric model into the rumor detection system and use cloud resources for real-time rumor detection and hazard metrics. Namely, we will develop the digital twin system application layer for social network rumor detection in cloud computing.

Authors' contributions

Hao Li, Wu Yang, Wei Wang, Huanran Wang contributed to the conception of the study; Hao Li performed the experiment; Hao Li, Wu Yang contributed significantly to analysis and manuscript preparation; Hao Li performed the data analyses and wrote the manuscript; Hao Li, Wu Yang, Wei Wang, Huanran Wang helped perform the analysis with constructive discussions.

Funding

This project was sponsored by National Key R&D Program of China (grant No. 2021YFB3101401), NSFC-Xinjiang Joint Fund Key Program (grant no.U2003206).

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Competing interests

The authors declare no competing interests.

Received: 23 November 2023 Accepted: 13 January 2024 Published online: 08 February 2024

References

- Bondielli A, Marcelloni FJIS (2019) A survey on fake news and rumour detection techniques 497:38–55
- Sun M et al (2023) Inconsistent Matters: A Knowledge-Guided Dual-Consistency Network for Multi-Modal Rumor Detection. IEEE Trans Knowl Data Eng 35(12):12736–12749. https://doi.org/10.1109/TKDE.2023.32755 86
- Hu X et al (2023) Mr2: A benchmark for multimodal retrieval-augmented rumor detection in social media. in Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23). Association for Computing Machinery, New York, NY, USA, p 2901–2912. https://doi.org/10.1145/3539618.3591896
- Lei Z, Chen Y, Lim MK (2021) Modelling and analysis of big data platform group adoption behaviour based on social network analysis. Technol Soc 65:101570. https://doi.org/10.1016/j.techsoc.2021.101570
- 5. Peng, K., et al., (2023) TOFDS: A Two-Stage Task Execution Method for Fake News in Digital Twin-Empowered Socio-Cyber World
- DiFonzo N (2013) Rumour research can douse digital wildfires. Nature 493(7431):135–135
- Sandor A (2020) The power of rumour(s) in international interventions: MINUSMA's management of Mali's rumour mill. Int Aff 96(4):913–934
- Greenspan RL, Loftus EF (2020) Pandemics and infodemics: Research on the effects of misinformation on memory. Hum Behav Emerg Technol 3(1):8–12. https://doi.org/10.1002/hbe2.228
- Kostkova P et al (2017) Who is Spreading Rumours about Vaccines? Influential User Impact Modelling in Social Networks, in Proceedings of the 2017 International Conference on Digital Health. Association for Computing Machinery, London, United Kingdom, pp 48–52
- 10. Zhou K et al (2019) Early Rumour Detection. Association for Computational Linguistics, Minneapolis, Minnesota

- Ma J et al (2015) 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. Association for Computing Machinery, Melbourne, Australia, pp 1751–1754
- 12. Zhang H et al (2020) Multimodal disentangled domain adaption for social media event rumor detection 23:4441–4454
- Zhang H et al (2019) Multi-modal Knowledge-aware Event Memory Network for Social Media Rumor Detection, in Proceedings of the 27th ACM International Conference on Multimedia. ssociation for Computing Machinery, Nice, France, 1942–1951
- Cambria E et al (2013) New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intell Syst 28(2):15–21
- ALDayel A, Magdy W (2021) Stance detection on social media: State of the art and trends. Inf Process Manage 58(4):102597. https://doi.org/10. 1016/j.ipm.2021.102597
- Tran T et al (2019) Misinformation Harms During Crises: When The Human And Machine Loops Interact. n 2019 IEEE International Conference on Big Data (Big Data)
- Wang P, Hu Y, Li Q (2020) The Trust-Building Process in the Social Media Environment of Rumour Spreading, in Companion Proceedings of the 2020 ACM International Conference on Supporting Group Work. Association for Computing Machinery, Sanibel Island, Florida, USA, pp 95–98
- Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter, in Proceedings of the 20th international conference on World wide web. ACM: Hyderabad, India, pp 675–684
- 19. Zhang Y et al (2021) Analysing rumours spreading considering selfpurification mechanism. Connect Sci 33(1):81–94
- 20. Nekovee M et al (2007) Theory of rumour spreading in complex social networks. Physica A 374(1):457–470
- 21. Liu X et al (2022) Weibo Spammer Detection Based On Social Network Digital Twin. in 2022 IEEE 2nd International Conference on Digital Twins and Parallel Intelligence (DTPI)
- Arisekola K, Madson K (2023) Digital twins for asset management: Social network analysis-based review. Autom Constr 150:104833
- 23. Zubiaga A et al (2018) Discourse-aware rumour stance classification in social media using sequential classifiers 54(2):273–290
- Hanselowski A et al (2018) A Retrospective Analysis of the Fake News Challenge Stance-Detection Task. Association for Computational Linguistics, Santa Fe, New Mexico, USA
- Ma J, Gao W, Wong K-F (2018) Detect Rumor and Stance Jointly by Neural Multi-task Learning. Companion of the the Web Conference 2018 on the Web Conference 2018 - WWW '18. https://doi.org/10.1145/3184558. 3188729
- Xuan K, Xia R (2019) Rumor stance classification via machine learning with text, user and propagation features. in IEEE 2019 International Conference on Data Mining Workshops (ICDMW), Beijing, China, p 560–566. https://doi.org/10.1109/ICDMW.2019.00085
- Pamungkas EW, Basile V, Patti V (2019) Stance Classification for Rumour Analysis in Twitter: Exploiting Affective Information and Conversation Structure. ArXiv.org. https://doi.org/10.48550/arXiv.1901.01911
- Hardalov M et al (2022) Few-Shot Cross-Lingual Stance Detection with Sentiment-Based Pre-training. Proceedings of the AAAI Conference on Artificial Intelligence 36(10):10729–10737
- 29. Godbole N, Srinivasaiah M, Skiena S (2007) Large-Scale Sentiment Analysis for News and Blogs. Icwsm 7(21):219–222
- 30. Bo P et al (2008) Opinion Mining and Sentiment Analysis. Found Trends Inf Retr 2(1–2):1–135
- Wang Z, Guo YJN (2020) Rumor events detection enhanced by encoding sentimental information into time series division and word representations 397:224–243
- Al-Ghadir Al, Azmi AM, Hussain A (2021) A novel approach to stance detection in social media tweets by fusing ranked lists and sentiments. Information Fusion 67:29–40
- Wang G, Tan L, Shang Z, Liu H (2022) Multimodal Dual Emotion with Fusion of Visual Sentiment for Rumor Detection. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.2204.11515
- Hossain MS, Muhammad G (2019) Emotion recognition using secure edge and cloud computing. Inf Sci 504:589–601

- Ghorbani M et al (2020) ConvLSTMConv network: a deep learning approach for sentiment analysis in cloud computing. Journal of Cloud Computing 9(1):16
- 36. Wang J et al (2021) Cloud-based intelligent self-diagnosis and department recommendation service using Chinese medical BERT. Journal of Cloud Computing 10(1):4
- Touvron H, Lavril T, Izacard G, Martinet X, Lachaux M-A, Lacroix T, Rozière B, Goyal N, Hambro E, Azhar F, Rodriguez A, Joulin A, Grave E, Lample G (2023) LLaMA: Open and Efficient Foundation Language Models. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.2302.13971
- Raffel C et al (2020) Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res. 21(1):140
- Ni J et al (2022) Sentence-T5: Scalable Sentence Encoders from Pretrained Text-to-Text Models. Association for Computational Linguistics, Dublin, Ireland
- Chung HW, Hou L, Longpre S, Zoph B, Tay Y, Fedus W, Li E, Wang X, Dehghani M, Brahma S, Webson A, Gu SS, Dai Z, Suzgun M, Chen X, Chowdhery A, Narang S, Mishra G, Yu AW, Zhao V (2022) Scaling Instruction-Finetuned Language Models. https://doi.org/10.48550/arxiv. 2210.11416
- Zhou H et al (2022) MDMN: Multi-task and Domain Adaptation based Multi-modal Network for early rumor detection. Expert Syst Appl 195:116517
- 42. Ma T et al (2021) A novel rumor detection algorithm based on entity recognition, sentence reconfiguration, and ordinary differential equation network. Neurocomputing 447:224–234
- Li Q, Huang Y, Jiang J (2015) Opinion index of social governance on Chinese network. Bull Chin Acad Sci 30(1):90–96. http://old2022.bulletin. cas.cn/zgkxyyk/ch/reader/view_abstract.aspx?file_no=20150112&flag=1
- 44. Hutto C, Gilbert E (2014) VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media 8(1):216–225
- Torshizi AS, Ghazikhani A (2019) Automatic Twitter Rumor Detection Based on LSTM Classifier. Springer International Publishing, Cham
- Alsaeedi A, Al-Sarem M (2020) Detecting Rumors on Social Media Based on a CNN Deep Learning Technique. Arab J Sci Eng 45(12):10813–10844. https://doi.org/10.1007/s13369-020-04839-2

Publisher's Note

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