JUSTC

Article

http://justc.ustc.edu.cn

Received: October 01, 2021; Accepted: December 14, 2021

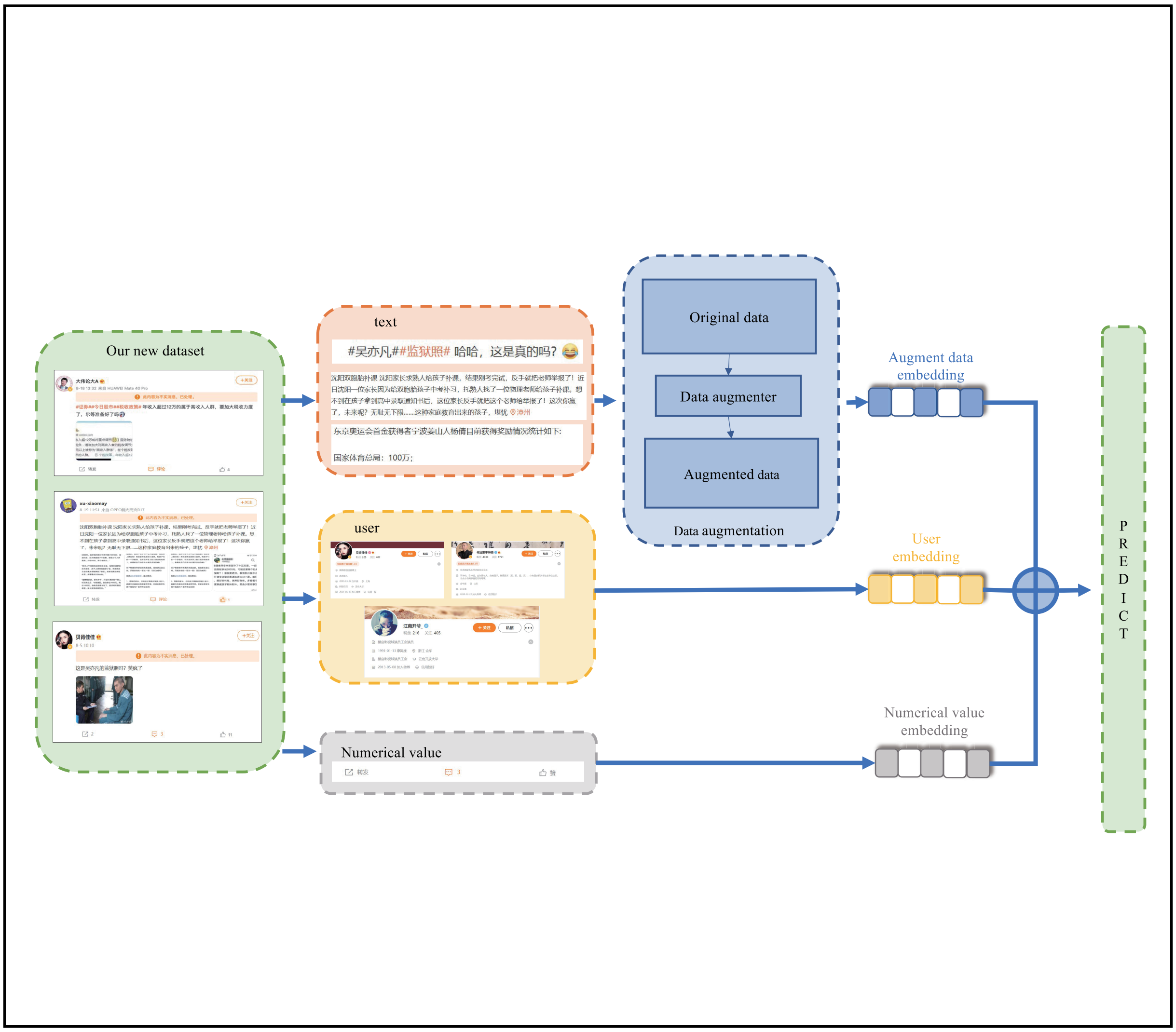
A data-driven model for social media fake news detection

Xin Chen, Shancheng Fang, Zhendong Mao ✉, and Yongdong Zhang

School of Information Science and Technology, University of Science of Technology of China, Hefei 230027, China

✉Correspondence: Zhendong Mao, E-mail: zdmao@ustc.edu.cn

Graphical abstract



#吴亦凡##监狱照# 哈哈，这是真的吗？😂

沈阳双胞胎补课 沈阳家长求熟人给孩子补课，结果刚考完试，反手就把老师举报了！近日沈阳一位家长因为给双胞胎孩子中考补习，托熟人找了一位物理老师给孩子补课。想不到在孩子拿到高中录取通知书后，这位家长反手就把这个老师给举报了！这次你赢了，未来呢？无耻无下限……这种家庭教育出来的孩子，堪忧 漳州

东京奥运会首金获得者宁波姜山人杨倩目前获得奖励情况统计如下:

国家体育总局：100万；

The overall framework of our fake news detection model.

Public summary

- We create a new Chinese social platform fake news dataset containing high quantity content and abundant information to facilitate research on fake news detection.
- Creatively introduce data augmentation in fake news detection research, and combine user attributes to better realize fake news detection.

Citation: Chen X, Fang S C, Mao Z D, et al. A data-driven model for social media fake news detection. *JUSTC*, **2022**, 52(3): 7. DOI: 10.52396/JUSTC-2021-0215

Article

http://justc.ustc.edu.cn

Received: October 01, 2021; Accepted: December 14, 2021

A data-driven model for social media fake news detection

Xin Chen, Shancheng Fang, Zhendong Mao ✉, and Yongdong Zhang

School of Information Science and Technology, University of Science of Technology of China, Hefei 230027, China

✉Correspondence: Zhendong Mao, E-mail: zdmao@ustc.edu.cn

Cite This: *JUSTC*, **2022**, 52(3): 7 (8pp)　　Read Online

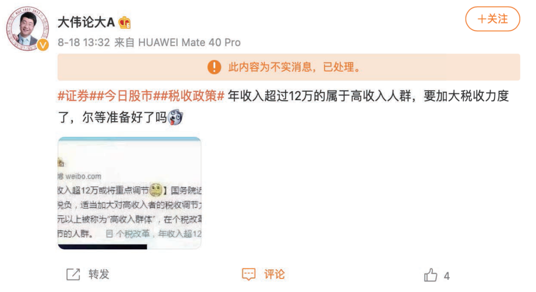
Abstract: The rapid development of social media leads to the spread of a large amount of false news, which not only affects people's daily life but also harms the credibility of social media platforms. Therefore, detecting Chinese fake news is a challenging and meaningful task. However, existing fake news datasets from Chinese social media platforms have a relatively small amount of data and data collection in this field is relatively old, thus being unable to meet the requirements of further research. In consideration of this background, we release a new Chinese Weibo Fake News dataset, which contains 26320 fake news data collected from Weibo. In addition, we propose a fake news detection model based on data augmentation that can effectively solve the problem of a lack of fake news, and we improve the generalization ability and robustness of the model. We conduct numerous experiments on our Chinese Weibo Fake News dataset and successfully deploy the model on the web page. The experimental performance proves the effectiveness of the proposed end-to-end model for detecting fake news on social media platforms.

Keywords: fake news detection; deep learning; machine learning

CLC number: TP391.1　　**Document code:** A

1 Introduction

In the era of the Internet, the channels through which people send and receive information are more abundant, which makes communication and exchange convenient. It is straightforward to share information and interact with one another. Online social media platforms (e.g., Weibo, Douyin, and Facebook) play a more important role in producing content and information propagation. However, fake news spreads spontaneously on these social media platforms, and some people with malicious intent use social media platforms to spread fake news virally[1−3]. Fig. 1 shows a piece of fake news about taxes on Weibo. The proliferation of fake news causes public panic, disrupts social order, affects public opinion, and manipulates the focus of the public; thus, fake news has become a significant destabilizing factor on social media[4]. For example, widespread fake news on social media has caused public panic during the ongoing Covid-19 crisis[5]. Therefore, the effective detection of fake news on social media has great significance for maintaining the stability of social life and cyberspace security.


大伟论大A

8-18 13:32 来自 HUAWEI Mate 40 Pro

此内容为不实消息，已处理。

#证券##今日股市##税收政策# 年收入超过12万的属于高收入人群，要加大税收力度了，尔等准备好了吗

博 weibo.com

收入超12万或将重点调节】国务院近

税负，适当加大对高收入者的税收调节

元以上被称为"高收入群体"，在个税改革

节的人群。 个税改革，年收入超12

Fig. 1. A piece of fake news about taxes on social media Weibo.

In general, there is no clear definition of fake news. The Merriam-Webster Online Dictionary states fake news as ' News reports that are intentionally false or misleading'[6]. Shu defined fake news as a verifiably false piece of information shared intentionally to mislead readers[7]. Ajao et al. defined it on online social media as ' any story circulated, shared, or propagated which cannot be authenticated'[8]. However, in academic research, fake news is typically defined as an unverified or unconfirmed message. In this study, we define fake news as misleading information that has proven to be false.

In recent years, academics and industrialists worldwide have paid more attention to the spread of fake news on social media platforms. To solve this problem, a large number of fake news analyses and fake news detection works on social media have appeared and achieved remarkable results. The identification method most easily proposed by researchers is the use of detection technology based on manual features[9−10]. On the one hand, with the development of artificial intelligence technology, recognition technology based on machine learning is proposed. It is a popular method to use classical text information representation methods such as n-gram and bag-of-words models[7], and then construct a supervised classifier[11]. Since the introduction of deep learning, there have been many applications of deep learning to mine higher-level feature representations for fake new detection[12−13]. They learn better text representations or sequence features from fake news. In addition, user characteristics and information dissemination structures are important features. There are a large number of malicious accounts on social networks that deliberately spread content containing misinformation to manipulate people's ideas and change people's decisions[14], and fake news detection methods focus on detecting social network users[15], such as using user attributes (e.g., language used by the account, geographic locations of the account, and authenticated identity)[16−17]. On the other hand, many researchers have focused on international social media platforms (Facebook, Flicker, etc.) while fewer researchers have focused on Chinese social media platforms. Although there are a few jobs geared towards Chinese social media platforms such as Weibo[18], they either analyze limited cases or a limited amount of data. It is difficult for models to sufficiently learn the characteristics of fake news to achieve reliable fake news detection results[12]. This greatly limits the progress and development of fake news analysis and detection and limits the comprehensive understanding of fake news on social media. Therefore, more data are required to enhance the discriminative power of the automatic fake news detection model.

Fake news detection has social significance, but there are some problems in practical applications. Manual fake news detection is undoubtedly the most reliable method for this purpose. For example, foreign social media anti-rumor websites require experts to analyze information and provide evidence to clarify whether the information is fake news. The domestic social media platform Weibo provides false information reporting functions. The annual report shows that the number of monthly reports on Weibo was as high as 127200, with a minimum of 74100. Since correctness of the news is completely determined by humans, it is highly dependent on the ability of the appraiser. The shortcomings of knowledge and the long detection cycle of fake news are very obvious; thus, this method has an explosive growth rate with the spread of information, and the scale of fake news increases exponentially and gradually fails to meet detection needs. It is challenging to quickly distinguish between users who maliciously report microblogs and who reasonably report questionable microblogs. Therefore, a large amount of data has also been used to test manpower requirements and manual review efficiency.

In this study, considering the current problems of fake news detection, we first propose a Chinese Weibo Fake News dataset that includes more than 20000 fake news data and more than 30000 real news data. Compared with previous Chinese fake news datasets, this significantly expands the data richness of the existing Chinese social network fake news dataset. The dataset also includes a wealth of user information and reporting information on fake news on social networking sites, including user reporting reasons, to facilitate further research. In addition, we evaluated the effectiveness of some current fake news detection methods on our dataset. Second, based on our dataset, we propose a novel method that combines contextualized representations from large pre-trained language models , such as BERT[19] or XLNet[20], with the machine learning model LightGBM[21] for fake news detection in Chinese social media. The contributions of this study are summarized as follows.

(i) We created a new Chinese social media platform fake news dataset containing a high quantity of content and abundant information to facilitate research on fake news detection.

(ii) We propose a framework based on post content and user attributes for fake news detection and evaluate the effectiveness of our dataset.

(iii) Our model can realize end-to-end fake news detection and achieve stable and accurate results, which can reduce the burden of manual reviews.

The rest of the paper is organized as follows: Section 2 provides a review of related works, Section 3 presents our created dataset, and Section 4 introduces our proposed method. Experiments are presented in Section 5. Finally, conclusions are presented in Section 6.

2 Related work

Some researchers make great contributions to the analysis and detection of fake news; thus, we mainly provide a brief review of related work from the following directions: data collection and fake news detection methods.

Dataset-related: Song et al.[22] collected large-scale social media fake news data, which uses Weibo as a research platform, and conducts a more comprehensive quantitative statistical semantic analysis of Chinese social media fake news. Ma et al.[12] proposed two datasets in 2016, and many subsequent fake news detection tasks were performed on these two datasets. Shu et al. proposed two datasets from PolitiFact and GossipCop from English websites in 2017[7, 23, 24]. Wang et al.[25] collected a dataset named the WeChat dataset from WeChat's Official Accounts, which is the largest instant messaging software in China.

Content-based: In recent years, many researchers have focused their attention on text content to determine the authenticity of information. Early research on fake news detection used linguistic features such as text length, word classes, and the percentage of pronouns[26, 27]. According to the analysis of the text characteristics of the post, they then used these characteristics to classify it as credible or untrustworthy; for example, the TFIDF and topic characteristics[28]. Malicious online users spread fake opinions by confusing language, writing styles[29], or sensational emotions[30]; therefore, some approaches have considered this in their work[31−33]. Chen et al.[34] utilized attention mechanisms based on a recurrent neural network (RNN) to learn text features for fake news recognition.

User-based: Another group of researchers believes that user analysis is a key part of fake news detection. Most users on social networks are ordinary people performing normal social activities, but there are a small number of malicious users deliberately creating fake news to affect the emotions of other users and achieve their own goals , such as enjoying attention from the public, triggering users' negative thoughts toward the country, and interfering with politics and other purposes. With such clear characteristics, there is a clear difference between malicious and normal users. Yang et al.[35] considers account-based features which include 'is verified,' 'has a description,' gender, avatar type, and name type. Shu et al.[36] analyzed explicit and implicit user profile features from social media platforms, where explicit characteristics refer to characteristics that can be directly obtained, such as the user's name and gender. Implicit features refer to information not directly obtained from user meta-information but can be inferred from other data, such as user historical tweets. The method proposed by Ma et al.[12] models a user response that

DOI: 10.52396/JUSTC-2021-0215

captures rich information in order to learn hidden user representations. Liu et al.[37] employed an incorporated RNN and convolutional neural network (CNN) to capture the representation characteristics of users based on time series, which can address early detection limitations. Qian et al.[38] proposed a two-level convolutional neural network (TCNN-URG) with a user response generator.

Propagation-based: Recently, some researchers have begun to exploit graph structures for fake news detection. The spread of fake news on social networks is similar to the spread of epidemics among the population. There are graph structures between users that are grouped by interests, similar opinions, and interactions with the news creator[24]. The diffusion patterns of online news are similar to the graph structure. Therefore, leveraging graph structures in fake news detection is a significant method for researchers. Refs.[13, 39–41] utilized the structural characteristics of message propagation by constructing propagation networks. Kai constructed publisher-news relation and user-news interaction networks to capture the potential relations among publishers, news pieces, and users[42]. Yuan et al.[43] captured the local and global relationships among all source tweets, retweets, and users by modeling a global heterogeneous graph. In contrast to previous ideas, Sampson et al.[44] combined the implicit links in conversation structures with the inherent network structure to increase the accuracy of fake news classification.

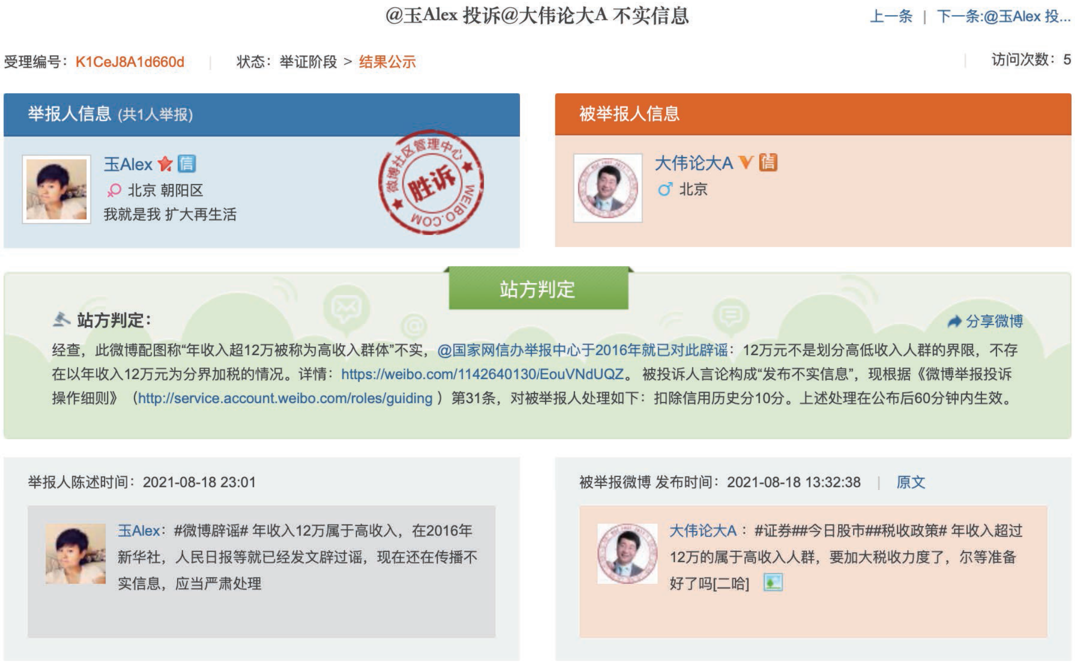
3 Dataset

In this section, we introduce the Chinese Weibo Fake News dataset in detail.

3.1 Data collection

Sina Weibo is the social media platform with the highest number of users in China. According to the latest Sina Weibo user survey report, the current daily active users of Sina Weibo have exceeded 200 million, which leads to massive amounts of information being posted on the platform every day. Fake news that misleads people mixed with life information is likely to have serious consequences. The public continues to pay attention to the Covid-19 crisis. In the early stage of the epidemic, much fake news was spread on the Weibo platform by users with ulterior motives. People's panic has caused extremely negative social effects. Therefore, Weibo officials have taken many measures to deal with similar situations. As of 2012, Weibo officials issued a series of management conventions and launched the Sina Weibo's official fake news busting service on the Weibo Community Management Center. The report processing hall is specifically used to handle users' reports of various false information and show the processing results publicly, such as deducting credit points, deleting Weibo posts, or banning users' accounts. In our dataset, we collected fake news microblogs from Weibo's official fake news busting service. Through Sina Weibo's official fake news busting service, we can obtain more official and authoritative fake news data, greatly avoiding errors in manual data annotation.

As shown in Fig. 2, we provide an example of the handling of fake news using Sina Weibo's official fake news busting service. Below, we introduce in detail the judgment of Sina Weibo's official fake news busting service on fake news. Sina Weibo's official fake news busting service report processing page displays the reported person and corresponding information. The reporting interface also provides the reporting time and the reasons provided by the reporting person. Furthermore, the interface displays the official judgment results, judgment basis, and processing method. Clicking the related hyperlink enters the user's homepage and microblog, which allows us to obtain important user information and details of the microblog.



@玉Alex 投诉@大伟论大A 不实信息

上一条 | 下一条:@玉Alex 投...

受理编号：K1CeJ8A1d660d | 状态：举证阶段 > 结果公示 | 访问次数：5

举报人信息 (共1人举报)

玉Alex
北京 朝阳区
我就是我 扩大再生活

微博社区管理中心 胜诉 WEIBO.COM

被举报人信息

大伟论大A
♂ 北京

站方判定

站方判定：

分享微博

经查，此微博配图称"年收入超12万被称为高收入群体"不实，@国家网信办举报中心于2016年就已对此辟谣：12万元不是划分高低收入人群的界限，不存在以年收入12万元为分界加税的情况。详情：https://weibo.com/1142640130/EouVNdUQZ。被投诉人言论构成"发布不实信息"，现根据《微博举报投诉操作细则》（http://service.account.weibo.com/roles/guiding ）第31条，对被举报人处理如下：扣除信用历史分10分。上述处理在公布后60分钟内生效。

举报人陈述时间：2021-08-18 23:01

玉Alex：#微博辟谣# 年收入12万属于高收入，在2016年新华社，人民日报等就已经发文辟过谣，现在还在传播不实信息，应当严肃处理

被举报微博 发布时间：2021-08-18 13:32:38 | 原文

大伟论大A：#证券##今日股市##税收政策# 年收入超过12万的属于高收入人群，要加大税收力度了，尔等准备好了吗[二哈]

Fig. 2. An example of Sina Weibo's official fake news busting service page.

We collected microblogs published by reported users on fake news busting services, from December 20, 2010, to December 01, 2020. Our dataset contains 26320 fake news items and 35,426 real news items.

3.2 Data analysis

The annual distribution of the number of fake news microblogs for our dataset is shown in Fig. 3. According to the official Weibo report, although fake news has increased yearly, most of the original microblogs are deleted due to being confirmed fake news on the platform.Therefore, such data do not appear in our data set. Otherwise, in our dataset, although some microblogs have similar themes and topics, their semantics may differ slightly. Considering the impact of different texts on the analysis, we did not exclude these microblogs. As far as possible, we consider various situations encountered in real-life fake news detection.

Each microblog in the Chinese Weibo Fake News dataset corresponds to a user. A total of 24181 users in our dataset posted fake news microblogs, of which only 4902 were officially certified. In contrast, the vast majority of fake news is from ordinary users. According to statistics on the number of followers, the majority accounted for less than 500 people. On the one hand, most ordinary users are incapable of discriminating against fake news; once the content of the fake news meets their daily interests, they can easily participate in its spread. On the other hand, there are a large number of virtual accounts on Weibo, and people with ulterior motives use these accounts to spread fake news. These accounts have never been used since posting fake news; therefore, there is no way to punish the users. The distribution of the number of followers of users posting fake news is shown in Fig. 4.

To show the influence of fake news, we use the sum of the number of reposted comments and likes of each microblog as an indicator to measure the influence of Weibo, and through statistical analysis of this indicator, we can understand the distribution of the influence of the rumor microblog. Owing to the differences in the topic, release time, and users, the influence of each microblog is different. The statistical information of this influence is shown in Fig. 5.

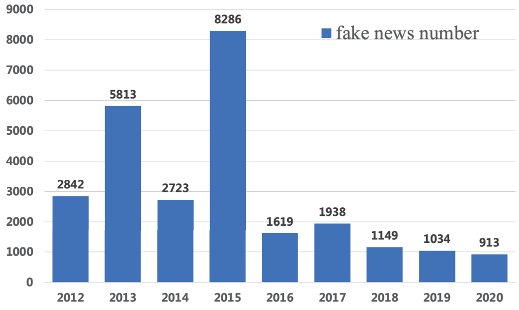


Fig. 3. Statistics of the annual number of fake news.

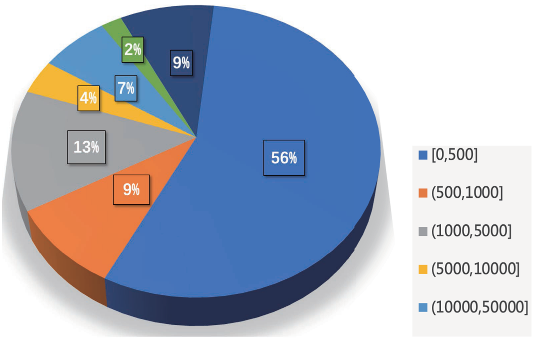


Fig. 4. Statistics of the number of followers of fake news accounts.

We find that the influence of most fake news is limited; microblogs with less than 500 influence accounts for 92% of the total. A few fake news microblogs have a large influence, which may cause serious consequences. Therefore, detecting fake news that may have a large influence on important information is a challenge that should be considered.

4 Method

In this section, we introduce a data-driven combination model to improve the fake news detection performance. Fig. 6 shows the proposed framework, which can be summarized into the following components: (i) The input consists of a variety of content of microblog and corresponding users' information. (ii) We use the augmenter to augment the data of limited fake news microblogs. (iii) The original data and augmented data contextualized representations and users' descriptions are obtained by Bert. (iv) These representations are concatenated and input to LightGBM for fake news detection. The details of the proposed framework are described below.

4.1 Data augmentation

In the fake news detection task, the imbalance of data and the small amount of fake news data are important factors that hinder development. In this study, we propose a method to generate new "fake news" samples to improve the performance of the model. For the original sample, $\boldsymbol{X} = \{x_1, x_2, \ldots, x_i\} \in \mathrm{R}^d$ in the training data, where i indicates the number of training data. We randomly generate a certain number of enhanced samples $\boldsymbol{X}_A = \{x_{1a\backslash 1b}, x_{2a\backslash 2b}, \ldots, x_{ia\backslash ib}\} \in \mathrm{R}^d$, where x_{1a} indicates the augmented data of x_1 and a and b represent different augmented methods, which are introduced as follows:

Word-level data augmentation: Based on the original data, we create new data that are similar to the original data by replacing synonyms, deleting unimportant words, and randomly inserting synonyms. We use the Jieba word segmentation tool to segment Chinese fake news text and then process the stop words in the text. On this basis, we randomly modify the words in the text with a ratio of 0.2, according to the

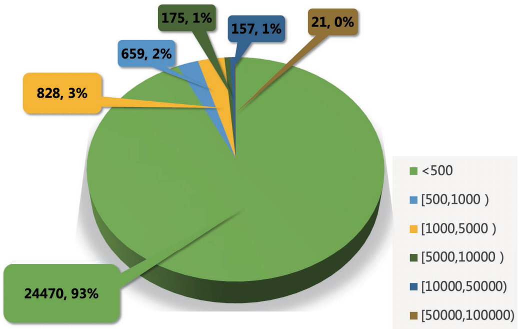


Fig. 5. Statistics of the influence of fake news.

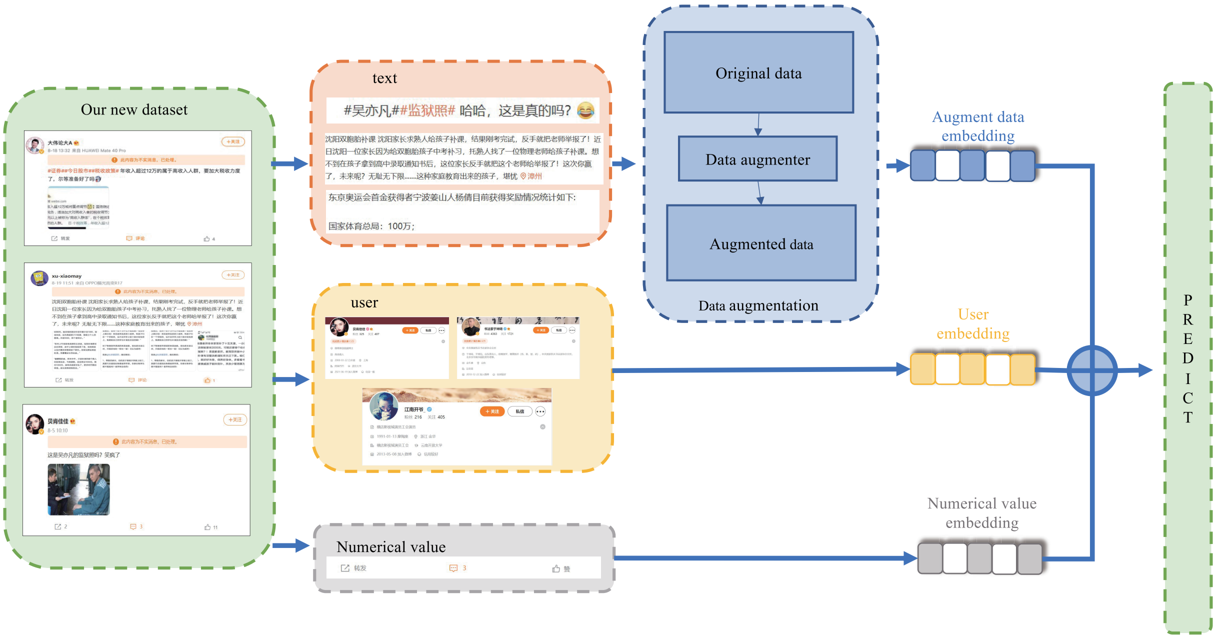

Our new dataset
text
#吴亦凡##监狱照# 哈哈，这是真的吗？
沈阳双胞胎补课 沈阳家长求熟人给孩子补课，结果刚考完试，反手就把老师举报了！近日沈阳一位家长因为给双胞胎孩子中考补习，托熟人找了一位物理老师给孩子补课。想不到在孩子拿到高中录取通知书后，这位家长反手就把这个老师给举报了！这次你赢了，未来呢？无耻无下限……这种家庭教育出来的孩子，堪忧 漳州
东京奥运会首金获得者宁波姜山人杨倩目前获得奖励情况统计如下:
国家体育总局：100万；
user
Numerical value
Original data
Data augmenter
Augmented data
Data augmentation
Augment data embedding
User embedding
Numerical value embedding
PREDICT

Fig. 6. The overall framework of our fake news detection model, which consists of the following steps：(i) The input consists of a variety of content of microblog and corresponding users’ information. (ii) We use the augmenter to augment the data of limited fake news microblogs. (iii)The original data and augmented data contextualized representations and users’ descriptions are obtained by Bert. (iv) These representations are concatenated and input to LightGBM for fake news detection.

method described above, and obtain an enhanced text that is different from the original text. We select more similar sentences by calculating their BLEU scores for the augmented sentences that are different from the original data obtained through different operations.

Sentence-level data augmentation: We use the free translation API provided by Google to translate the original data into other languages and then translate it back to Chinese. Owing to the different logical orders of various languages, this method can obtain new data with consistent semantics that compare with the original data.

Through the above method, we obtain different enhanced texts and select the last two enhanced texts by calculating the BLEU score.

4.2 Feature representation learning

Each microblog contains rich text information, and we obtain useful object representations using Latent Dirichlet Allocation (LDA) for fake news classification. LDA is an unsupervised probability generation technology that can be used to identify hidden subject information in text. The model assumes that "the article selects a topic with a probability and selects a word from this topic with a certain probability," where the text-to-topics and topics-to-words follow a polynomial distribution, which is described as follows:

$$P_{(x_1,x_2,\ldots,x_k;n,p_1,p_2,\ldots,p_k)}=\frac{n!}{x_1!\ldots x_k!}p_1^{x_1}\ldots p_k^{x_k} \tag{1}$$

By observing our fake news data, we find that most of the topics in fake news are related to social and political topics. Through LDA, we obtain hidden topic information in the text, which is represented as $f_{\rm LDA}$.

The meaning of the text cannot be fully understood using only the LDA method, and we use the Bert model as the backbone to obtain all data features that contain the original text and augmented text. Bert use a transformer as its main framework. The transformer abandons the traditional CNN and uses the self-attention mechanism to better solve the problem of long-term dependence on text and training speed. The attention mechanism enables the model to focus on all the input information that is important for the target word, which can greatly capture global information and use parallel training to increase the speed such that the model effect is greatly improved. In addition, the change in the attention weight matrix allows us to understand the role of different words in the task more intuitively, which has strong interpretability. Bert carry out sufficient self-supervised learning based on a massive corpus to learn a good feature representation for words and has a strong feature representation ability. Bert can also use expectations related to specific tasks for fine-tuning to improve the performance of the final model. We use $f_{\rm Bert}$ to denote the text representations learned by Bert. Finally, we concatenate these into textual representations, denoted as follows:

$$f_{\rm text}=[f_{\rm LDA}||f_{\rm Bert}] \tag{2}$$

where || represents the splicing operation.

Research has shown that on social media platforms, user behavior characteristics have a certain degree of group aggregation. For example, people with the same hobbies and educational backgrounds are more likely to gather together. Because a social structure is composed of many users and their activities, the importance of users in social networks is obvious. In social networks, a certain number of naval accounts are dedicated to spreading fake news. However, there are many official authoritative accounts on social media platforms and their published content is strictly reviewed. The content of such accounts has a relatively high credibility. Therefore, the user attributes of social accounts occupy a higher weight in fake news detection. User attributes are composed of the text information described by the user, number of users' fans, number of followers, and whether they are authenticated. In this study, we utilize the pre-trained Bert model to process user descriptions, which are denoted as $f_{\text{description}}$. We use other numerical features directly, which is denoted as $f_{\text{numerical}}$. Because we use a classification model based on a tree structure, this model is trained through feature splitting. Numerical scaling does not affect the location of the split point or the structure of the tree model. We concatenate our learned user text features and numerical features to obtain the final representation of the user, which is expressed as follows:

$$f_{\text{user}} = [f_{\text{description}} \| f_{\text{numerical}}] \tag{3}$$

After obtaining the augmented text features and user features, we use the LightGBM [21] model to obtain the final fake news prediction results. LightGBM is a framework that implements the gradient boosting decision tree (GBDT) [45] algorithm. The main idea of the GBDT is to use weak classifiers (decision trees) to iteratively train and obtain the optimal model, which is challenging to overfit and fast in training. At the same time, when deploying the model to our online fake news detection system, we found that some microblogs did not provide user information. Therefore, we train the two models, and the final output was the weighted effect of the two models.

5 Experiments

In this section, we discuss the evaluation indicators of the fake news detection algorithms. In addition, we introduce the effect of the baseline experiment on fake news detection and validate the proposed model on this dataset.

5.1 Dataset

In this study, we utilize the Chinese Weibo Fake News dataset to assess the proposed model. The details of our dataset are presented in Section 3. Table 1 lists the dataset statistics.

5.2 Evaluation metrics

In the experiment, the trained model is evaluated using standard classification performance metrics: accuracy, precision, recall (sensitivity), and F1-score. First, we introduce the concept of a confusion matrix, as shown in Table 2.

Based on the confusion matrix, we define the evaluation metrics mentioned above as follows:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{4}$$

Table 1. Statistics of the Chinese Datasets.

	Our dataset
#News	61746
#Fake news	26320
#True news	35426
#Fake user	24181

Table 2. Confusion Matrix of evaluation metrics.

	Positive	Negative
True	True Positive(TP)	True Negative(TN)
False	False Positive(FP)	False Negative(FN)

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{6}$$

$$\text{F1-score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{7}$$

In general, a higher result means better performance.

5.3 Baseline approaches

The state-of-the-art fake news detection methods compared with our framework are as follows:

Naive Bayes: The Naive Bayes method is a classification method based on Bayes' theorem and the assumption of the independence of characteristic conditions. We employ the count vectorizer and TFIDF vectorizer in Naive Bayes.

SVM [46]: The support vector machine (SVM) is a two-classification model. The basic model is a linear classifier with the largest interval defined in the feature space. The kernel technique used by the SVM essentially makes it a nonlinear classifier to model microblog content.

LSTM: We employ the pre-trained Chinese word vector [47] to represent the microblog and then build a Bi-LSTM network structure. Finally, these representations are input into a fully connected layer to make predictions. The parameters of Bi-LSTM are set to 64 and 32. Built on top of the Bi-LSTM, a fully connected layer outputs the probability that the news is fake.

GRU [12]: The gate recurrent unit (GRU) uses an RNN to model the text and extract the features; it performs similarly to LSTM, but it is easy to train and can greatly improve training efficiency.

CNN [48]: The CNN model uses a convolutional neural network to learn fake news representations for each microblog. The extracted microblog representations are finally passed through a fully connected layer with a softmax function to predict the fake news result.

5.4 Performance comparison

The performances of our proposed framework and other excellent algorithms on our dataset are presented in Table 3. For the parameter selection of the classification model, we set the maximum depth of the decision tree to five, learning rate to 0.12, bagging fraction to 0.8, and feature fraction to 0.8. The proposed framework achieves the best performance.

Our model achieves an F1-score of 0.903, with an accuracy of 0.890 for the test data. It achieves the best results compared with the other methods in the fake news mask. The prediction performance of the Naive Bayes method is much lower than that of other methods, regardless of whether it is using the CountVectorizer word-embedding vector method or

Table 3. Comparison with state-of-the-art methods on our dataset.

Method	Accuracy	Precision	Recall	F1
NB-Count	0.729	0.735	0.727	0.733
NB-TFIDF	0.730	0.732	0.729	0.730
SVM	0.792	0.784	0.802	0.793
LSTM	0.838	0.849	0.890	0.869
GRU	0.843	0.850	0.898	0.873
CNN	0.858	0.874	0.886	0.880
Ours	0.890	0.897	0.912	0.899

the TFIDF word-embedding vector method. The main reason for this is that Bayes classification is more sensitive to feature forms. Compared with Naive Bayes classifiers, the performance of SVM is improved, its generalization ability is better, and it is not as sensitive to data forms. However, compared with the deep learning model, the performance of early machine learning is still much worse. Deep learning methods can directly learn advanced features from data, automatically extract features, and integrate feature learning into model construction, thereby overcoming the shortcomings of traditional feature engineering tasks. However, deep learning methods require a large amount of data training to achieve a better understanding.

Moreover, none of the above methods consider the important role of user information in fake news detection. We consider user information, and through a weighted combination of the fake news detection model for plain text and the model containing user information, our method has higher accuracy and flexibility in practical applications.

6 Conclusions

In this study, we propose an end-to-end data augmentation combination framework that uses data augmentation technology and the joint modeling of microblog text information and user attribute characteristics for fake news detection. Considering the problem of fake news data with little time and strong time characteristics, the model is enhanced to improve the generalization ability and robustness of the fake news detection task utilizing data augmentation and makes full use of the microblog text characteristics and user attribute characteristics to improve the performance of the model. Finally, in practical applications, a weighted combination of plain text and increased user characteristics is used to obtain a more stable prediction performance. Our proposed framework preforms rich experiments on the Sina Weibo data and achieves the best performance, which indicates the effectiveness of the method for fake news detection. In the future, our model could be extended to other tasks. In addition, we will continue to explore additional detection methods for fake news.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (WK3480000008, WK3480000010, WK2100000023).

Biographies

Xin Chen is currently pursuing a master’s degree at the University of Science and Technology of China, Hefei, China. Her research focuses mainly on rumor detection in social media and cross-modal understanding.

Zhendong Mao received his PhD degree in computer application technology from the Institute of Computing Technology, Chinese Academy of Sciences, in 2014. He was an assistant professor with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing, from 2014 to 2018. He is currently a professor at the School of Cyberspace Science and Technology, University of Science and Technology of China, Hefei, China. His research interests include computer vision, natural language processing and cross-modal understanding.

References

[1] Zhou Y Q. A Study of the Rumors in the Internet of Contemporary China. Beijing: The Commercial Press, 2012.

[2] Allport G W, Postman L. The Psychology of Rumor. New York: Henry Holt, 1947.

[3] Peterson W A, Gist N P. Rumor and public opinion. *American Journal of Sociology,* **1951**, *57*: 159–167.

[4] Liu F, Burton-Jones A, Xu D. Rumors on social media in disasters: Extending transmission to retransmission. In: Proceeding of the 19th Pacific Asia Conference on Information Systems (PACIS 2014), Chengdu, China. 2014: 49.

[5] Tasnim S, Hossain M M, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. *Journal of Preventive Medicine and Public Health,* **2020**, *53* (3): 171–174.

[6] Fake news. In: 7 Words from Political Scandals. https://www.merriam-webster.com/words-at-play/political-scandal-words/fake-news.

[7] Shu K, Sliva A, Wang S, et al. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter,* **2017**, *19* (1): 22–36.

[8] Ajao O, Bhowmik D, Zargari S. Sentiment aware fake news detection on online social networks. In: ICASSP 2019: 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 2507–2511.

[9] Wang W Y. "Liar, liar pants on fire": A new benchmark dataset for fake news detection. https://arxiv.org/abs/1705.00648.

[10] Pérez-Rosas V, Kleinberg B, Lefevre A, et al. Automatic detection of fake news. https://arxiv.org/abs/1708.07104.

[11] Cha M, Gao W, Li C T. Detecting fake news in social media: An Asia-Pacific perspective. *Communications of the ACM,* **2020**, *63* (4): 68–71.

[12] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16). New York: IJCAI, 2016: 3818–3824.

[13] Ma J, Gao W, Wong K F. Rumor detection on Twitter with tree-

structured recursive neural networks. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, 2018: 1980–1989.

[14] Davis C A, Varol O, Ferrara E, et al. BotOrNot: A system to evaluate social bots. In: Proceedings of the 25th International Conference Companion on World Wide Web. Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2016: 273–274.

[15] Yang X, Lyu Y, Tian T, et al. Rumor detection on social media with graph structured adversarial learning. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20) . Virtual, Japan: IJCAI, 2020: 1417–1423.

[16] Ferrara E, Varol O, Davis C, et al. The rise of social bots. *Communications of the ACM,* **2016**, *59* (7): 96–104.

[17] Zhao J, Cao N, Wen Z, et al. # FluxFlow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics,* **2014**, *20* (12): 1773–1782.

[18] Sun S, Liu H, He J, et al. Detecting event rumors on Sina Weibo automatically. In: Web Technologies and Applications. APWeb 2013. Berlin: Springer, 2013.

[19] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. https://arxiv.org/abs/1810.04805.

[20] Yang Z, Dai Z, Yang Y, et al. XLNet: Generalized autoregressive pretraining for language understanding. In: Advances in Neural Information Processing Systems 32 (NeurIPS 2019). Red Hook, NY: Curran Associates Inc., 2019, 32: 5753–5763.

[21] Ke G, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems 30 (NIPS 2017). Red Hook, NY: Curran Associates Inc., 2017, 30: 3149–3157.

[22] Song C, Yang C, Chen H, et al. CED: Credible early detection of social media rumors. *IEEE Transactions on Knowledge and Data Engineering,* **2021**, *33* (8): 3035–3047.

[23] Shu K, Mahudeswaran D, Wang S, et al. FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data,* **2020**, *8* (3): 171–188.

[24] Shu K, Wang S, Liu H. Exploiting tri-relationship for fake news detection. https://arxiv.org/abs/1712.07709.

[25] Wang Y, Yang W, Ma F, et al. Weak supervision for fake news detection via reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence,* **2020**, *34* (01): 516–523.

[26] Rubin V L, Conroy N, Chen Y, et al. Fake news or truth? Using satirical cues to detect potentially misleading news. In: Proceedings of the Second Workshop on Computational Approaches to Deception Detection. San Diego, California: Association for Computational Linguistics, 2016: 7–17.

[27] Mihalcea R, Strapparava C. The lie detector: Explorations in the automatic recognition of deceptive language. In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. Suntec, Singapore: Association for Computational Linguistics, 2009: 309–312.

[28] Castillo C, Mendoza M, Poblete B. Information credibility on Twitter. In: Proceedings of the 20th International Conference on World Wide Web. New York: ACM, 2011: 675–684.

[29] Potthast M, Kiesel J, Reinartz K, et al. A stylometric inquiry into hyperpartisan and fake news. https://arxiv.org/abs/1702.05638.

[30] Zhang X, Cao J, Li X, et al. Exploiting emotions for fake news detection on social media. https://arxiv.org/abs/1903.01728.

[31] Przybyla P. Capturing the style of fake news. *Proceedings of the AAAI Conference on Artificial Intelligence,* **2020**, *34* (01): 490–497.

[32] Popat K. Assessing the credibility of claims on the web. In: Proceedings of the 26th International Conference on World Wide Web Companion. Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2017: 735–739.

[33] Potthast M, Kiesel J, Reinartz K, et al. A stylometric inquiry into hyperpartisan and fake news. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, 2018: 231–240.

[34] Chen T, Li X, Yin H, et al. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. In: Trends and Applications in Knowledge Discovery and Data Mining. Cham, Switzerland: Springer, 2018: 40–52.

[35] Yang F, Liu Y, Yu X, et al. Automatic detection of rumor on Sina Weibo. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. New York: ACM, 2012: 1-7.

[36] Shu K, Zhou X, Wang S, et al. The role of user profiles for fake news detection. In: 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. New York: ACM, 2019: 436–439.

[37] Liu Y, Wu Y F B. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018: 354–361.

[38] Qian F, Gong C, Sharma K, et al. Neural user response generator: Fake news detection with collective user intelligence. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18). Stockholm, Sweden: IJCAI, 2018: 3834–3840.

[39] Wu L, Liu H. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. New York: ACM, 2018: 637–645.

[40] Monti F, Frasca F, Eynard D, et al. Fake news detection on social media using geometric deep learning. https://arxiv.org/abs/1902.06673.

[41] Shu K, Mahudeswaran D, Wang S, et al. Hierarchical propagation networks for fake news detection: Investigation and exploitation. *Proceedings of the International AAAI Conference on Web and Social Media,* **2020**, *14*: 626–637.

[42] Shu K, Wang S, Liu H. Beyond news contents: The role of social context for fake news detection. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. New York: ACM, 2019: 312–320.

[43] Yuan C, Ma Q, Zhou W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection. In: 2019 IEEE International Conference on Data Mining (ICDM). IEEE, 2019: 796–805.

[44] Sampson J, Morstatter F, Wu L, et al. Leveraging the implicit structure within social media for emergent rumor detection. In: Proceedings of the 25th ACM international on Conference on Information and Knowledge Management. New York: ACM, 2016: 2377–2382.

[45] Friedman J H. Greedy function approximation: A gradient boosting machine. *Annals of Statistics,* **2001**, *29* (5): 1189–1232.

[46] Hassan S, Rafi M, Shaikh M S. Comparing SVM and naïve Bayes classifiers for text categorization with Wikitology as knowledge enrichment. In: 2011 IEEE 14th International Multitopic Conference. IEEE, 2011: 31–34.

[47] Li S, Zhao Z, Hu R, et al. Analogical reasoning on Chinese morphological and semantic relations. https://arxiv.org/abs/1805.06504.

[48] Yu F, Liu Q, Wu S, et al. A convolutional approach for misinformation identification. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17). Melbourne, Australia: IJCAI, 2017: 3901–3907.

DOI: 10.52396/JUSTC-2021-0215