International Journal for Housing Science and Its Applications
Publish August 5, 2025. Volume 46, Issue 3 Pages 4203-4212

https://doi.org/10.70517/ijhsa463359

Study on the Sentiment Dynamics of Public Opinion after the Sichuan Luding MS 6.8 Earthquake Based on BERT Sentiment Analysis

Guoyong Pan[1,2], Ye Ren[1,2], Haiying Yu[1,2], Xia Bo[1,2] and Xiuqing Song[1,2,*]
[1] Shanghai Earthquake Agency, Shanghai, 200062, China
[2] Shanghai Sheshan National Geophysical Observatory, Shanghai, 200062, China
Corresponding authors: (e-mail: dzzysdyj@126.com).

Abstract A large amount of information related to earthquakes is released and widely disseminated on the Internet, but there may also be some misleading views, which may generate a large public opinion risk if emergency countermeasures are not quickly formulated. In this paper, multi-user web crawler technology is used to obtain the text data of public opinion after the Sichuan Luding MS 6.8 earthquake, often the crawled text data contain disturbing information, and the data preprocessing work is completed through the steps of word splitting, de-duplication and labeling. After that, the text is transformed into word vectors that can be read directly by computers using the BERT model. The word vectors are put into the LSTM model for training, so as to realize the dynamic monitoring of public opinion sentiment of Sichuan Luding MS 6.8 earthquake. With the dual support of dataset and evaluation indexes, the model of this paper is evaluated and analyzed. Within 72 hours after the Sichuan Luding MS 6.8 earthquake, the information exposure rises to 532647420660 items, which attracts extensive attention from the society, and with the evolution of time, people's concern about the earthquake public opinion will gradually decline, which comprehensively outlines the change of public opinion sentiment dynamics. In addition, in the intelligent detection of Sichuan Luding earthquake public opinion sentiment dynamics, this paper's model detects accurately up to 93.42%, much better than the CNN model, which indicates that this paper's model is able to guide users to the correct public opinion.

Index Terms BERT, LSTM, Luding, earthquake public opinion

I. Introduction

The development and popularization of Internet technology has gradually changed people's way of life, the life of food, clothing, housing and transportation have been inseparable from the Internet, people are also more inclined to express their thoughts and feelings on the network. Especially when an earthquake occurs, a large amount of public opinion information containing disaster information will be generated on the Internet within a short period of time, and the collection of disaster information in the first time is the key to the disposal of earthquake emergency response [1], [2]. In addition, the huge number of Internet users express their personal opinions and attitudes through self-media and other channels, and a large amount of information related to earthquakes is released and widely disseminated among media platforms, which is gathered into a huge amount of data, which contains information about earthquake public opinion and disaster conditions [3]-[6]. However, there may also be some misleading opinions, which may generate a large public opinion risk if emergency countermeasures are not quickly formulated [7], [8]. Therefore, it is necessary to fully excavate and analyze relevant information to quickly and accurately control the trend of earthquake public opinion, to help relevant departments to monitor, guide and dispose of public opinion, and then to maintain social stability [9]-[11].

At the same time, according to the increasing requirements of refinement and timeliness of earthquake emergency service products, real-time, accurate and comprehensive mastery of disaster information is carried out [12], [13]. BERT is a bidirectional representational pre-trained language representation model, whose structure is mainly made up of encoders stacked in the Transformer model [14]. Designing a deep learning earthquake public opinion sentiment dynamic analysis model based on BERT can provide reference for earthquake online public opinion monitoring by virtue of the model's efficient feature extraction capability [15], [16].

In this project, text vector processing techniques, text data preprocessing, BERT word vector techniques, and LSTM models are first discussed in depth to facilitate the implementation of the subsequent research work. Multi-user web crawling technology is utilized to obtain Sichuan Luding MS6.8 earthquake public opinion data from microblogs, and some data preprocessing operations are done for the characteristics of the earthquake microblog comment data, and a high quality and large scale dataset applicable to the model of this paper is finally constructed. Comprehensive BERT word vector technology, LSTM model, complete the construction of Sichuan Luding MS6.8 earthquake public opinion sentiment dynamic detection model, and evaluate and analyze the model of this paper.

II. Sichuan Luding Earthquake Public Opinion Sentiment Discussion

II. A. Mathematical models

II. A. 1) Natural language processing techniques

Generally speaking, the process of natural language processing is roughly divided into data acquisition, data preprocessing, text representation, feature extraction (model training) and result evaluation, and the flow chart is shown in Figure 1. Data acquisition generally uses web crawler technology or local direct import and other ways to obtain data text. Data preprocessing mainly transforms the cluttered unstructured data into structured data and performs operations such as de-weighting, word splitting, and de-deactivating words [17]. Text representation is to transform the text into a vector or matrix form that can be understood by the computer. Feature extraction is the automatic extraction of semantic features from the text using machine learning or deep learning algorithm models. And result evaluation is to objectively analyze the output results of the model through a series of evaluation indexes, and then assess the goodness of the model. Among them, text representation and feature extraction are the key to determine the effectiveness of the model. In general, natural language processing is a constantly evolving field, and this paper aims to apply natural language processing to the earthquake field, transforming the determination of the Sichuan Luding earthquake public opinion into a text categorization problem in natural language processing, and providing new ideas for the management of earthquake public opinion. The main techniques in each process of natural language processing will be explained in detail next.

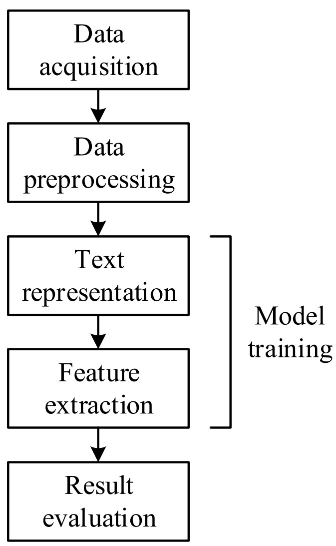


Figure 1: Flowchart of Natural Language processing

II. A. 2) Data pre-processing

Data preprocessing is one of the important parts of the natural language processing process. Generally, the data we collect through web crawlers and other technologies contain some unstructured, useless and repetitive information, such as blank information and missing information, which may have some impact on the training of the model, so we need to transform these data into structured data, and in addition, it also contains the steps of word splitting, de-duplication and labeling.

II. A. 3) Text representation

The BERT model is shown in Fig. 2. Text representation is the basis of natural language processing, which is the process of transforming text into a form that can be understood by computers, generally in the form of vectors or matrices, also known as word vector technology [18]. Text representation determines how the model extracts and utilizes information from the original text, so a good text representation is the cornerstone for improving the performance of the algorithmic model. There are mainly two forms of text representation, discrete and distributed, discrete representation is represented by One-hot coding, and the representative word embedding models of distributed representation are Word2Vec model and BERT model.Although Word2vec effectively solves the problem of high sparsity and high dimensionality of the One-hot word vectors, and successfully captures the richer semantic information, however it also There are certain limitations, the word vectors generated by Word2Vec are static, which only consider the contextual environment in which the word is located, but ignore the order of occurrence of the word in the sentence or text. In addition, Word2Vec generates a single, fixed vector representation for each word, which cannot effectively distinguish the different meanings of the same word in different contexts. Therefore the concept of dynamic word vectors has been proposed, which utilizes a multi-layer, bi-directional Transformer encoder to learn the textual representation, and by introducing a self-attention mechanism that allows it to simultaneously consider both pre- and post-contextual information about a word at any time, which enables BERT to better capture the complex relationship between words and contexts, and to generate a dynamic, contextually relevant vector representation for each word . In this way, even for polysemous words, BERT can generate different vectors according to their specific position in the sentence and the surrounding context, thus more accurately reflecting their actual meanings.

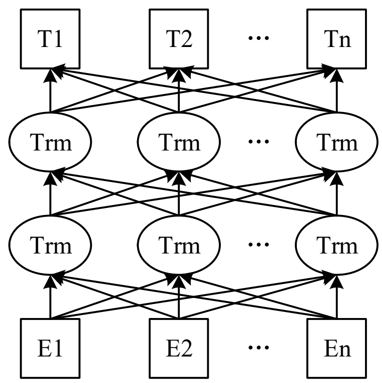


Figure 2: BERT model structure diagram

II. A. 4) Feature extraction

Feature extraction is an indispensable key link in natural language processing. With the rapid advancement of deep learning model algorithms in the field of natural language processing, we often leverage these deep learning model algorithms to deeply identify and extract semantic features in textual data, thus laying a solid foundation for more efficiently accomplishing natural language processing tasks. In this subsection, we will explore some classic deep learning models, including convolutional neural networks, recurrent neural networks, long and short-term memory networks, gate recursive unit networks, and FastText models. Each of these models has unique advantages and features that provide us with richer tools and means to process and understand natural language.

The recurrent mechanism of RNN will face the problem of gradient vanishing or gradient explosion when processing long sequences of text, in order to solve this problem, researchers proposed the Long Short-Term Memory Network (LSTM).The LSTM is an improved RNN model, the LSTM manages the flow of information by introducing a unique gating mechanism, which realizes the selective forgetting and retention of the information and effectively prevents gradient vanishing or explosion [19], [20].The LSTM gating mechanism mainly consists of forgetting gate, input gate and output gate its structure is shown in Fig. 3.

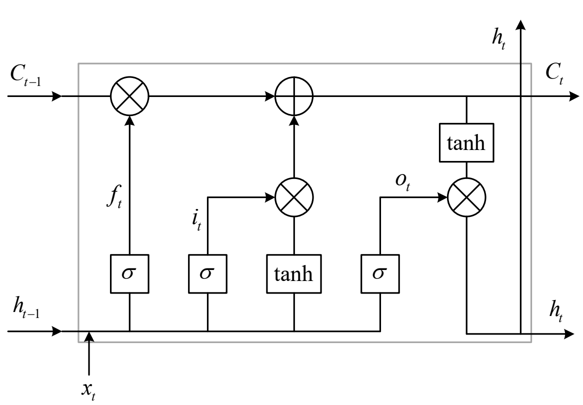


Figure 3: LSTM model structure diagram

The first gated oblivion gate f_t will determine how much information in the current cell state C_t is retained in the cell state C_t in the previous moment by using the inputs x_t at the current t-moment and the outputs h_{t-1} at the previous moment, which is calculated as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

where σ denotes the sigmoid function, W_f is the weights of the forgetting gate, and b_f is the bias of the forgetting gate.The sigmoid function converts the memory cells of the previous moment and the input of the current moment into a value between 0 and 1, which denotes the probability of forgetting, and the forgetting gate then forgets, based on this probability, the last moment of the part of the information in the memory cell is forgotten.

The second stage is the selective memory stage, where the input gate i_t decides how much of the input x_t of the current moment can be saved into the cell state C_t of the current moment, also by means of a sigmoid function. This is shown in the following equation. The candidate memory unit $\tilde{C}_t$ normalizes its content to -1 to 1 after going through the tanh layer, denoting the fraction of new information retained. Finally the current moment cell state C_t is updated by combining the forgetting gates of the first stage.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

The third stage is the output stage, LSTM is mainly through the output h_{t-1} from the previous moment and the input x_t at the current moment to determine the output gate o_t, and then through the activation function tanh layer, and ultimately get the output of the current moment, the formula is shown below.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \times \tanh(C_t) \tag{6}$$

II. B.Earthquake Public Opinion Commentary Dataset Construction

This chapter mainly explains the source of the microblog comment dataset for the Sichuan Luding earthquake event and the preprocessing done to the dataset, introduces the principle of the crawler design and how to crawl the comment data from microblogs, describes the data characteristics of the microblog comment of the Sichuan Luding earthquake and the sentiment annotation for the microblog comment data of the earthquake, and so on.

II. B. 1) Earthquake twitter comment data collection

The source of the dataset constructed in this thesis is the comment data of earthquake disaster events in Sina Weibo, and the principle of designing the crawler is based on the requests framework, which searches for the relevant blog posts through the information of the three elements of the earthquake as the keywords after the earthquake, and then crawls the data and saves it in the database through the multi-user crawling technique. In Sina Weibo, the information of official political media is often both correct and fast, and the comment information under these blog posts is usually the most useful information. With the magnitude of different earthquakes, there will be subsequent reports of casualties and secondary disasters, and these comment information is also the most timely and useful public opinion information. Therefore, this paper adopts a multi-user crawling mechanism to crawl earthquake microblog comments, and Table 1 shows examples of users crawled in this study. This time, the data collection of 2022 Sichuan Luding earthquake microblog comments is the data source for model training, and a series of data information in the microblogs of these earthquakes within a week after the occurrence of the earthquake is collected, which mainly includes the user information, time information, the body of the microblog comments, the number of likes, and other key public opinion information.

Table 1: Crawl the Weibo user

Weibo user	Weibo user	Weibo user
@China Earthquake Networks Center	@Southern Daily	@Shock length
@Rapid report from China Earthquake Networks Center	@Headline News	@Sichuan Fire Department
@National Emergency Broadcasting	@Chinese meteorology enthusiast	@(Some unofficial bloggers)
@People's Daily	@China Fire Protection	@Sichuan Daily
@China Newsweek	@Sichuan.com	@CCTV News

II. B. 2) Preprocessing of the Earthquake Microblog Commentary Dataset

Earthquake Weibo comments have their own data characteristics, as shown in Table 2 below. After the data is collected, the seismic microblog comment data is preprocessed, and the crawled microblog comment data is first cleaned, for example, the data will contain characters such as "@XXX" and "#XX", so this paper needs to regularize the text and remove the non-Chinese characters in the comment data, as well as @, # Relevant content after characters, and spaces. According to the characteristics of the comment text of the earthquake microblog, the stop word list of the earthquake microblog comments was constructed, and the stop words were removed from the comment text after the earthquake. The emoji expression in the comment text also needs to be processed, because the expression will also contain a certain emotional tendency in the text, so it cannot be deleted directly, that is, this paper uses the emoji switch library to convert the emoji expression into the corresponding Chinese text, and together with the text information to form a dataset. Finally, the preliminarily cleaned earthquake Weibo comment dataset was formed, and a total of 180111 post-earthquake Weibo comment data were collected after the above operations. Then, the data cleaned earthquake microblog comment dataset was labeled, and the Chinese emotional vocabulary ontology database compiled by the learning team of Dalian University of Technology was used as the source of emotional annotation used in the model, and "happy" and "good" were defined as positive emotions (positive), "anger", "sadness", "fear" and "surprise" were defined as negative emotions (negative), and the rest were (neutral). Finally, the annotation of these data was carried out with three categories of emotions. Due to the relatively small number of earthquake microblogs, especially strong earthquakes, and the dimensions of earthquake microblog comment data that can be collected are not wide, resulting in fewer earthquake microblog comment features that can be extracted during subsequent model training, this paper uses NLP data augmentation technology EDA to enhance the data through synonym substitution, random insertion, random exchange, etc., so as to expand the amount of post-earthquake Weibo comment data. Finally, a dataset of 385283 earthquake microblog comments was obtained after cleaning, annotation, and enhanced preprocessing.

Table 2: Characteristics of Weibo comment data on earthquakes

Text form	Data volume	Timeliness	Herd effect
The comments after the earthquake are short texts Most of them have flexible forms of expression Diversity.	The magnitude of the earthquake and the number of people involved Follow-up reports on casualties, secondary disasters, etc. The Dao is closely related	Comments on Weibo are real-time. After official government media publish blog posts, users will post comments. At various stages after the earthquake, the number of comments will increase along with the users' attention.	The so-called "herd effect" means that when Weibo users read blog posts about earthquakes, they will see the comments of other users and relate to them. Comments of interest can also be replied to, liked, etc., to express one's own opinions in real time.

II. C.General research framework

Aiming at the Sichuan Luding earthquake microblog comment text has the problems of containing specialized comments, few extractable features and multiple meanings of the word, this thesis proposes a BERT-LSTM Sichuan Luding earthquake public opinion sentiment analysis model. Since the deep learning model can learn the features in the text from the training data with good generalization ability, for the problems of polysemy and few extractable features in the earthquake microblog comment data, taking into account that the BERT model is able to represent the Sichuan Luding Earthquake microblog comment text data with word vectors that contain contextual semantic information, the LSTM model provides a more comprehensive identification and extraction of the features of the data, and the LSTM model to identify and extract features more comprehensively, and to assign weights to important parts of the comment text to realize the demand for sentiment classification of earthquake microblog comments. Therefore, this paper uses the BERT model as the text vectorization representation layer, transforms the seismic microblog comment text into the earthquake microblog comment word vector containing contextual semantic information, uses the LSTM network as the feature extraction layer, and inputs the word vector sequence into the LSTM layer to extract the text features of the seismic microblog comment, and then strengthens the weight of the preliminary extracted features, and finally obtains the emotional tendency result value of the earthquake microblog comment. The BERT-LSTM model proposed in this paper makes full use of the contextual information of the context and the sentiment information related to the earthquake public opinion, which solves to some extent the problems of the BERT model in the sentiment analysis of the earthquake microblog comments such as the problem of the extractable features are few, and the problem of the polysemy of words due to the failure to consider the syntactic dependency between words.

III. Exploration and Analysis of Public Opinion on Sichuan Luding Earthquake

III. A. Overview of the earthquake situation

According to the official determination of the China Earthquake Network, a magnitude 6.8 earthquake occurred at 12:52 on 2022-09-05 in Luding County, Ganzi Prefecture, Sichuan Province, at a depth of 16km. after the earthquake, the China Earthquake Administration (CEA) quickly organized and carried out emergency response in accordance with the plan, continued to carry out seismic monitoring and tracking as well as trend analysis and judgment, and quickly dispatched an on-site team of more than 160 people, comprised of relevant provincial earthquake departments and operational units, to the quake zone to carry out emergency response work. The China Earthquake Administration (CEA) quickly organized and carried out emergency response in accordance with the plan, continued to monitor and track the earthquake situation, analyzed and assessed the trend, and quickly dispatched an on-site team consisting of relevant provincial seismological departments and operational units, totaling more than 160 people, to the quake zone to carry out emergency response. According to the on-site investigation, the earthquake triggered many and extremely serious geological disasters, causing a large number of casualties, damage to buildings and infrastructure, and disruption of roads, communications, water supply and power supply and other lifelines in many places. According to the seismic intensity map released by the Ministry of Emergency Management on 2022-09-11, the highest intensity of the Sichuan Luding earthquake was IX degree (9 degrees), and the long axis of the isoseismic line was oriented to the north and west, with the long axis of 195 km and the short axis of 112 km, and the area of the VI degree (6 degrees) zone and above was 19,089 km², which involves a total of 12 counties (municipalities and districts) in 3 cities and states in Sichuan Province, with 82 townships and towns (streets). As of 2022-09-11 17:00, the earthquake has killed 93 people, including 55 in Ganzi Prefecture and 38 in Ya'an City.

III. B. Exploratory Analysis of Sichuan Luding Earthquake Public Opinion Dataset

III. B. 1) Analysis of the Number and Trend of Microblog Public Opinion Releases after the Earthquake

The number and trend of microblog public opinion releases after the earthquake are shown in Figure 4. Within 72h after the earthquake, 1254911 microblog messages related to this destructive earthquake event were collected through Sina Weibo, with 931653 retweets, 673009 comments, 14130630938 likes, and 532647420660 exposures. As a sudden public security incident, this 6.8 magnitude earthquake event in Luding County, Ganzi Prefecture, Sichuan Province, has attracted widespread attention from the society.

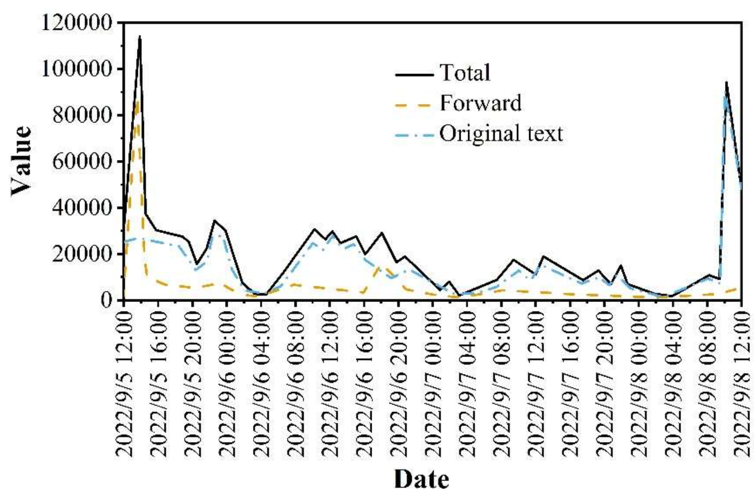


Figure 4: The number and trend of Weibo public opinion releases after the earthquake

The trend of the number of original information released within 72h after the earthquake is shown in Figure 5, and the average dissemination rate of microblog opinion information for the whole earthquake event is 17,248.77 items/h. The dissemination peaked at 2022-09-05 14:30, at 115,158 items. At 2022-09-05 14:30, about 2.5h after the earthquake, the number of original information published on Weibo public opinion showed a rapid decline, but still maintained a high degree of social concern in terms of the overall number and trend of publication.

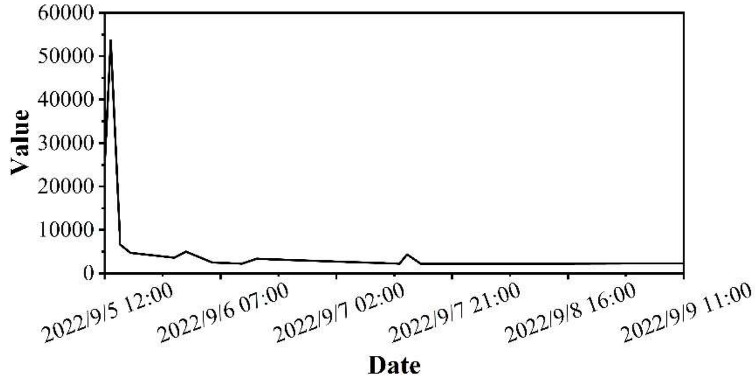


Figure 5: Original information release within 72 hours after the earthquake

III. B. 2) Statistical Analysis of Word Frequency of Microblog Public Opinion

The collected microblog opinion information is text-cleaned to filter out the retweeted microblog data without any information, and the statistical analysis of microblog opinion word frequency is shown in Figure 6. The information related to reflecting the earthquake situation is extracted by artificial intelligence natural language processing technology, and the hot word cloud is generated by using the word division tool. The top 100 hot words in frequency of occurrence are selected to visualize the disaster information in the public opinion of the 6.8 magnitude earthquake in Luding County, Ganzi Prefecture, Sichuan Province in terms of space and size, which can intuitively reflect the first impression of the public opinion information of this destructive earthquake event.

Figure 6: Statistical analysis of Weibo Public Opinion Word Frequency

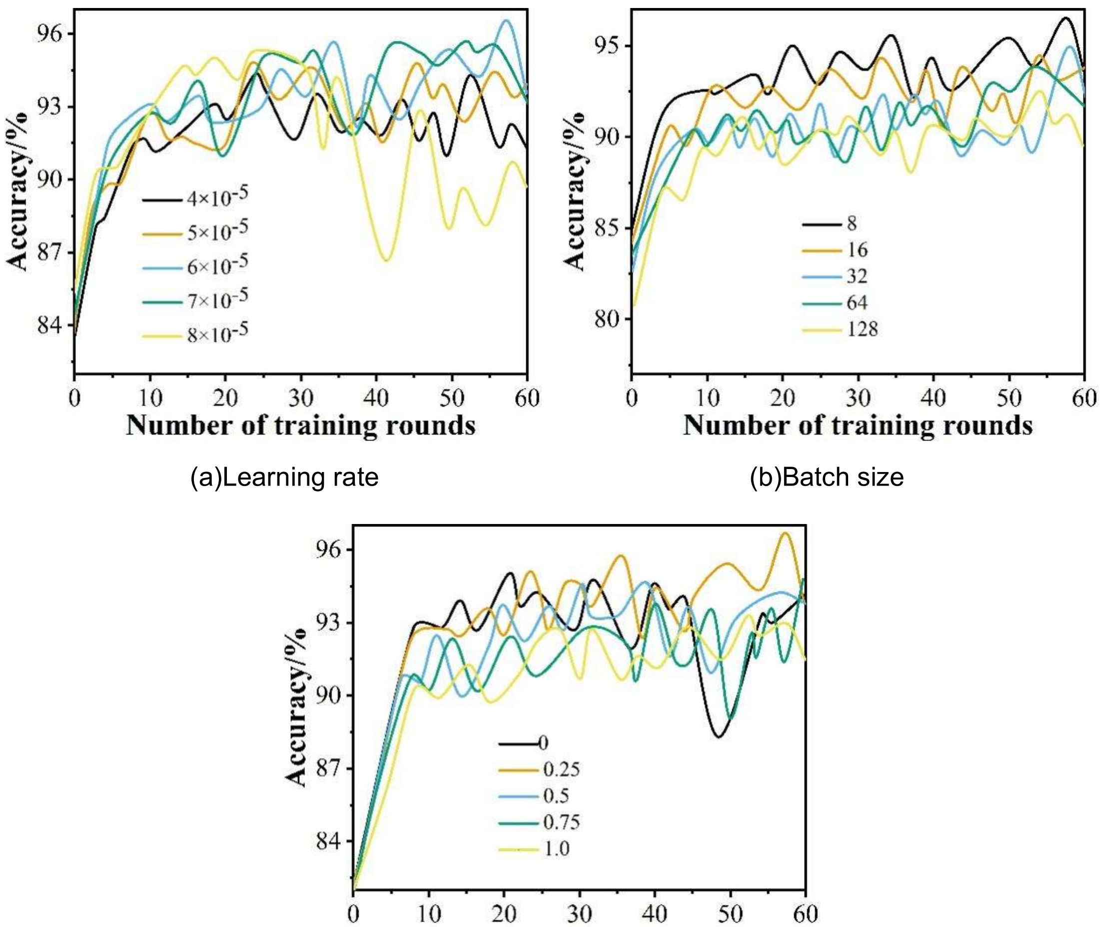
III. C. Experimental environment and model parameters

III. C. 1) Experimental environment

We use the base version of Google's open source BERT with 768 hidden layers, and select the open source pre-trained model “bert-base-chinese” of Hugging Face. The computer used for the experiment is Windows 8 with Intel(R) Core(TM) i7-10750H processor and 32 GB of RAM, and the model used for the experiment is the PyTorch deep learning framework, cu116 driver, and 1650 graphics card.

III. C. 2) Model parameters

In this paper, the output of the last layer of the BERT model is used as the feature input of the fully connected layer classifier, while the discard layer is utilized to prevent overfitting. The model training is done in a fine-tuned way, using Adam optimizer and binary cross-entropy loss function, with gradient trimming set to 1 (to prevent gradient explosion), and the maximum length of the input is set to 420 (each tweet is limited to 140 characters, and a Chinese character or punctuation accounts for 3 bytes in the UTF-8 encoding). In the course of the fine-tuning process, in order to find the optimal parameter combinations, the parameters such as the initial learning rate, the batch size, and the discard ratio are respectively compared and experimented, and each model parameter is trained for 60 rounds, and the parameters with the highest correct rate on the training set are selected. Comparison experiments were conducted, each model parameter was trained for 60 rounds, and the parameter with the highest correct rate on the training set was selected as the optimal value, and the performance comparison of different model parameters is shown in Fig. 7. As can be seen in Fig. 7(a), the learning rate is chosen to be 5 values such as 4×10-5, 5×10-5, 6×10-5, 7×10-5, and 8×10-5 (with a fixed batch size of 8 and a discard ratio of 0.25), the model accuracy decreases rapidly when the learning rate is greater than 7×10-5, and the model performance is optimal when the learning rate is 6×10-5. From Fig. 7(b), it can be seen that the batch size is taken as 5 values such as 8, 16, 32, 64 and 128 (with a fixed learning rate of 6 × 10-5 and a discard ratio of 0.25), and due to the memory constraints, the maximum length of the inputs is tuned down to 180 and 135 when the batch size is 64 and 128, respectively. The smaller the batch size, the higher the accuracy, so take a batch size of 8. From Fig. 7(c), it can be seen that the discard ratio is taken as five values such as 0, 0.25, 0.5, 0.75 and 1.0 (fixed learning rate is 6×10-5 and batch size is 8), and the model accuracy is highest when the discard ratio is 0.25. Based on the experimental comparison, the learning rate parameter of the model is taken as 6×10-5, the batch size parameter is taken as 8 and the discard ratio parameter is taken as 0.25.



(a)Learning rate

(b)Batch size

(c)Discard ratio

Figure 7: Performance comparison of different model parameters

III. D. Model evaluation analysis

III. D. 1) Evaluation indicators

Word vectors for the test set are generated using BERT as well. The word vectors are fed into the sentiment analysis model trained in the previous step to predict the sentiment labels of the generated test samples. The sentiment labels of the generated test set samples are compared with the manual labeling to evaluate the effectiveness of the model. In order to further carry out the evaluation of the model effect, this paper also selects the traditional CNN for experiments, the model features also use word embedding, the length of the convolution kernel of the onedimensional convolutional layer is 3, a total of 265. The evaluation metrics include Precision, Recall and the reconciliation metric F1-score of the two.By counting the number of positive and negative sentiments in the data set, the TP, FN, FP and TN are derived in Table 3.The formula for calculating the Precision, Recall and F1-score are as follows:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{8}$$

$$F1-score = 2\frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{9}$$

Table 3: Evaluation index parameters

	It is predicted to be positive emotion	It is predicted to be negative emotion
It is actually positive emotion	TP	FN
It is actually negative emotion	FP	TN

III. D. 2) Analysis of results

The comparison of emotion recognition effect is shown in Table 4, and the accuracy rate of LSTM is as high as 93.42% in the Sichuan Luding earthquake public opinion emotion recognition problem. LSTM is higher than traditional classical CNN in precision rate, recall rate and F1-score.0.2, 0.4, 0.6, 0.8 and 1.0 of the overall data are selected as the training data set for testing, and the results are shown in Fig. 8.LSTM has reached more than 90% precision rate when the training data is only 40%, and the approximation speed is faster. The experiment shows that LSTM has better effect and performance in the sentiment analysis of earthquake public opinion and can recognize the sentiment of the captured earthquake public opinion. The sensitivity of the public opinion classified as negative sentiment is compared with the sensitive thesaurus, and if it exceeds the minimum threshold value of 75%, it is judged to be a public opinion with harm, and the system will start the early warning response, send warning messages or emails to notify the relevant departments for network intervention. The fusion model based on BERT and LSTM has a higher accuracy rate in emotion recognition of Sichuan Luding earthquake public opinion. Earthquake rumors have a strong emotional tendency, and the earthquake public opinion emotion recognition model constructed in this paper provides strong support for the accurate identification of subsequent rumors.

Table 4: Comparison of emotion recognition effects

Index	CNN	LSTM
Precision	0.8867	0.9288
Recall	0.9084	0.9396
F1-score	0.8942	0.9342

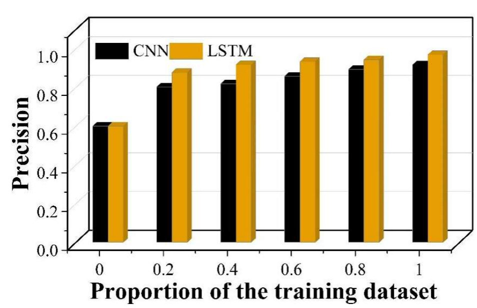


Figure 8: Performance comparison between LSTM and CNN

IV. Conclusion

In this paper, starting from the Sichuan Luding earthquake tremor, we crawl the data and save it in the database through multi-user crawling technology, and all the collected data contain some unstructured, useless and repetitive information. These information may have some impact on the subsequent research results, so we need to transform these data into structured data, in addition to containing the steps of word splitting, de-duplication and labeling. With the help of the BERT model, the text is transformed into word vectors that can be understood by computers, and on this basis, the LSTM model is used to more comprehensively identify and extract the features of the data as well as to assign weights to the important parts of the text of the comments, so as to realize the dynamic analysis of the sentiment of the earthquake microblog comments. It is found that the number of microblog retweets reaches an astonishing 931,653 within 72h after the Sichuan Luding earthquake, and the highest peak of the Sichuan Luding earthquake communication value is found at 2022-09-05 14:30, which fully expresses the dynamic change of the public opinion of the Sichuan Luding earthquake. Finally, the model parameters are set and the model of this paper is evaluated and analyzed. In the Sichuan Luding earthquake public opinion sentiment recognition problem, the accuracy rate of this paper's model is as high as 93.42%, which is significantly higher than that of the CNN model, indicating that the fusion model of BERT and LSTM has a much higher accuracy rate in Sichuan Luding earthquake public opinion sentiment recognition, thus providing technical support for governmental public opinion management.

Funding

This research was supported by Shanghai Science and Technology Program Project (Project Number:23DZ1203902) and the Shanghai Sheshan National Geophysical Observatory (Project Number:2024SSY11).

References

[1] Wang, C., Ye, Y., Qiu, Y., Li, C., & Du, M. (2024). Evolution and spatiotemporal analysis of earthquake public opinion based on social media data. Earthquake Science, 37(5), 387-406.
[2] He, W., Yuan, Q., & Li, N. (2023). Research on the characteristics of internet public opinion and public sentiment after the sichuan earthquake based on the perspective of weibo. Applied Sciences, 13(3), 1335.
[3] Chen, X. (2023). A Review on the Analysis of Public Opinion Regarding Natural Disasters. Advances in Engineering Technology Research, 7(1), 240-240.
[4] Liu, J., Zhu, S., Wang, Z., & Chen, S. (2024). The evolution of online public opinion on earthquakes: a system dynamics approach. Humanities and Social Sciences Communications, 11(1), 1-18.
[5] Li, S., & Sun, X. (2023). Application of public emotion feature extraction algorithm based on social media communication in public opinion analysis of natural disasters. PeerJ Computer Science, 9, e1417.
[6] Yang, T., Xie, J., Li, G., Mou, N., Li, Z., Tian, C., & Zhao, J. (2019). Social media big data mining and spatio-temporal analysis on public emotions for disaster mitigation. ISPRS International Journal of Geo-Information, 8(1), 29.
[7] Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. Sentiment analysis and ontology engineering: An environment of computational intelligence, 313-340.
[8] Pan, G., Ren, Y., Yu, H., & Song, X. (2025). Modeling and Visualization of Public Opinion Sentiment Evolution after Sichuan Luding MS6. 8 Earthquake Based on LSTM Neural Networks. J. COMBIN. MATH. COMBIN. COMPUT, 127, 9515-9535.
[9] Ying, L. I. U., Zhiming, L. I. U., & Haixia, J. I. N. (2023). A Study on Public Opinion Strategy of New Media Platforms of Earthquake in Emergency Responses. North China Earthquake Sciences, 41(2), 32-36.
[10] Chen, Y., Li, Y., Wang, Z., Quintero, A. J., Yang, C., & Ji, W. (2022). Rapid perception of public opinion in emergency events through social media. Natural hazards review, 23(2), 04021066.
[11] Yuan, Q., Wang, S., & Li, N. (2024). Research on emotional tendency of earthquake disaster based on E-Trans model: take the topic of “Sichuan Earthquake” on microblog as an example. Natural Hazards, 120(6), 5057-5074.
[12] Ragini, J. R., Anand, P. R., & Bhaskar, V. (2018). Big data analytics for disaster response and recovery through sentiment analysis. International Journal of Information Management, 42, 13-24.
[13] Huang, X., Jin, H., & Zhang, Y. (2019). Risk assessment of earthquake network public opinion based on global search BP neural network. PloS one, 14(3), e0212839.
[14] Wan, B., Wu, P., Yeo, C. K., & Li, G. (2024). Emotion-cognitive reasoning integrated BERT for sentiment analysis of online public opinions on emergencies. Information Processing & Management, 61(2), 103609.
[15] Li, Y. (2023). BERT for sentiment analysis in the era of epidemic. Applied and Computational Engineering, 6, 1217-1226.
[16] Li, R., Zhao, L., Xie, Z., Ji, C., Mo, J., Yang, Z., & Feng, Y. (2025). Mining and analyzing the evolution of public opinion in extreme disaster events from social media: A case study of the 2022 yingde flood in china. Natural Hazards Review, 26(1), 05024015.
[17] Jingmin Luan,Qiyang Li,Yuyan Qiu & Weican Liu. (2025). Ensemble learning unlocking point load forecasting accuracy: A novel framework based on two-stage data preprocessing and improved multi-objective optimisation strategy.Computers and Electrical Engineering,124(PA),110282-110282.
[18] Sheng Hua Xiong,Zhi Hong Wang,Zhen Song Chen,Gang Li & Hao Zhang. (2025). Text classification of public online messages in civil aviation: A N-BM²5 weighted word vectors method. Information Sciences,704,121956-121956.
[19] Jun Wang,Xiulai Wang & Airong Yu. (2025). Tackling misinformation in mobile social networks a BERT-LSTM approach for enhancing digital literacy. Scientific Reports,15(1),1118-1118.
[20] Janani Muralitharan & Chandrasekar Arumugam. (2024). Privacy BERT-LSTM: a novel NLP algorithm for sensitive information detection in textual documents. Neural Computing and Applications,36(25),15439-15454.