Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230611

BERT for sentiment analysis in the era of epidemic

Yanshu Li

School of Computer Science & Technology, Soochow University, Suzhou, 215000, China

2009404142@stu.suda.edu.cn

Abstract. With the continuous progress of Internet technology, the network platform has gradually entered everyone's life, providing a platform for ordinary people to express their ideas. Since the occurrence of COVID-19, monitoring and analyzing public opinion on the Internet platform has become more practical. Through timely monitoring and analysis, it is of great practical significance for the relevant departments to analyze and control sentiment information and stabilize and guide public sentiment. Therefore, it is essential and of practical significance to select a suitable model for classifying and analyzing public opinion on the Internet platform. This paper reviews the development of word vector technology from the perspective of technology development and then lead to the more advanced Bidirectional Encoder Representations from Transformers (BERT) model with great significance. On this basis, this paper fine-tunes the pre-trained Bert model. It applies the transfer learning strategy to analyzing the public sentiment of the occurrence of COVID-19 during the recent epidemic in Shanghai based on Sina Weibo data. In addition, tests are conducted to compare the model with the previous models. The experimental results show that the Bert model has significant advantages over the traditional model in character vector encoding and feature extraction.

Keywords: BERT, sentiment analysis, transformer, epidemic.

1. Introduction

Google researchers released a large-sized pre-trained language model based on a bidirectional transformer, Bidirectional Encoder Representation from Transformers (BERT), at the end of 2018. The fine-tuned model, based on pre-training, proved its powerful and advanced performance in the NLP field. Subsequently, the team disclosed the model's official code and pre-trained model. The remarkable thing is that the classification of public sentiment is a research hotspot in this field and once caused a surge of research. In the early stage, Zhu et al. proposed to employ a vector space model to classify Chinese text automatically [1]. Wang et al. developed a dictionary classification model based on word vectors [2]. Zhang et al. proposed using the text similarity method for text separation [3]. These studies can guide the early stage of research in this field. However, they are primarily based on statistical laws, can not be adjusted according to the changes of specific datasets, and thus can not meet the current requirements of public sentiment text analysis. After the rise of neural network technology, many studies on text classification using neural networks emerged. Kim et al. used the pre-trained word vector as input to realize text emotion classification with a Convolutional Neural Network (CNN) [4]. Kalchbrenner et al. introduced the concept of a dynamic Convolutional Neural Network (DCNN). They applied it to learn the features of sentences, finally achieving excellent results [5]. Sun et al. designed a GRU-attention classification model by connecting the attention layer based on the Gated Recurrent Unit and proved through experiments that the classification performance of this model is improved compared with that of the support vector machine (SVM) and GRU network alone [6]. The application of neural networks significantly improved the performance of text classification.

Nevertheless, the structure of such neural networks is relatively simple, and their ability for feature extraction and coding is weak. It is not easy to achieve ideal results in text classification tasks involving public sentiment analysis, which have complex semantics and need to consider the context. With further research advancement, more networks with more complex structures continued to emerge. Yue Liu et al. proposed to use CNN to extract the feature representation of phrase sequence, using N-LSTM (NestedLong Short-Term Memory) to learn the feature representation of text while introducing an attention mechanism to highlight key phrases to optimize feature extraction [7]. This processing enables the model to capture the sentence's local features and the context's semantic information and highlight the impact of each part of the input text on the text category. Compared with the baseline models, its performance on the three public datasets improved significantly. XinHui Liu et al. proposed the method of Multi-head Attention combined with CapsuleNet and bidirectional long-term and short-term memory network (BiLSTM) [8]. With this method, they successfully improved the F1 value of the model on Reuters-21578 and AAPD data sets to 89.82% and 67.48%, respectively.

Considering the above models for text classification, it can be found that if a more scientific and perfect model can be proposed, it can be fine-tuned according to its specific situation when dealing with a particular task. On the premise of ensuring the model's performance, it can adapt to the needs of the task rapidly, saving many research resources and improving labor efficiency. Therefore, this paper selects the Bert model for application research, analyzes its design principle, and uses it to classify public opinion texts during the epidemic. This paper fine-tunes the pre-trained BERT model to better apply it to text classification tasks. Therefore, this paper first looks at the structure and peculiarities in closer detail. This model does not apply a brand-new depth network or put forward a new algorithm at the technical level. Instead, it has improved and integrated various deep learning and natural language processing technologies in recent years.

This paper evaluates the performance of the BERT model for text sentiment classification. It compares it with other baseline models to explore the superiority of the BERT model. The goal is to summarize the relevance between public sentiment and social events according to the analysis results of the model.

2. Background

COVID-19 first broke out at the end of 2019. Up to now, sporadic cases still exist in various places, significantly affecting people's everyday lives. In the whole process of epidemic prevention and control, the Internet plays a significant role. It is the main channel for people to understand epidemic prevention and control measures. At the same time, people also constantly show their views on the Internet. The characteristics of disseminating online public sentiment are topic intensive, multilateral evolution, and derivative variety. Multiple subjects exchange opinions, discuss, and debate closely around relevant hotspots when communicating online and offline [9]. Therefore, carrying out targeted research on Internet public sentiment has a tremendously positive effect on maintaining social stability, pacifying people's emotions, and ensuring the regular operation of Internet life and real production life after the occurrence of unexpected social events. In recent years, platforms in the form of Internet forums have developed rapidly, and Sina Weibo (Weibo in short) is a prominent representative. With much information, Weibo provides valuable samples for detecting and analyzing online public sentiment [9]. Blog posts highly liked and forwarded on Weibo will cause a resonance effect. When some individuals think this viewpoint is reasonable, they tend to make the same remarks or opinions, thus affecting the behavior of a larger group [10]. The blogs which obtain many interactions on Weibo can reflect the views of netizens in a concentrated way. They contain solid personal emotional traits, reflecting the emotional tendencies of different groups [11]. Li et al. found through research that the emotional tendency of comments in great demand plays a guiding role [12]. Suppose a popular blog post shows a

DOI: 10.54254/2755-2721/6/20230611

positive emotional attitude. In that case, the feedback form in the following comments will probably be positive, and vice versa. Furthermore, various pieces of evidence show a specific relationship between the online news people are exposed to and the onset of anxiety and depression [13]. This paper introduces the epidemic situation and public opinion in Shanghai in the first half of 2022 as the research background in the next section.

2.1. Overview of the epidemic in Shanghai

The data of the epidemic in Shanghai were collected by official media, and the specific data are shown in Figure 1&Figure 2. The situation was stable from March 1 to March 22, and the daily increase was controlled within 1000 cases. Since March 23, the increase in new cases has been apparent. March 29 was the first turning point of the epidemic, and the number of new cases began to decline. However, since April 1, the number of new cases has increased dramatically. Since April 4, the number of cases has increased daily to 10000. On April 10, April 12, and April 16, there was a brief turning point, and then it continued to climb. On April 22, after the daily increase reached 23300 cases, an apparent downward trend began to appear. On April 28, the number of cases increased slightly, showing a downward trend. On April 30, the daily increase dropped to less than 10000 cases. Then it kept falling; by May 29, it had dropped to lessthan 100 cases.

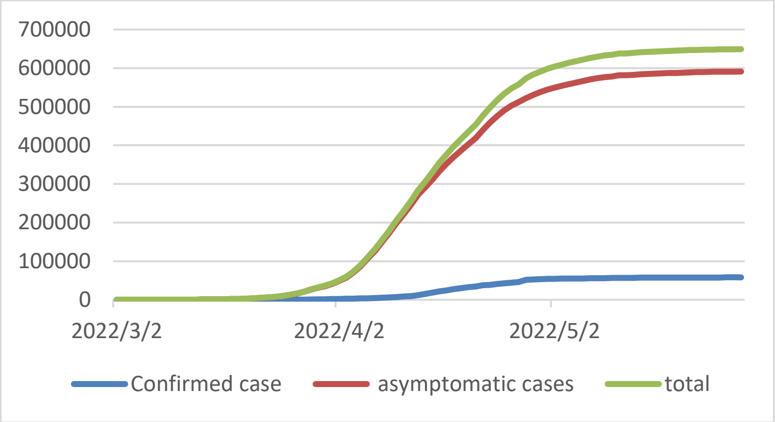

Figure 1. Total cases.

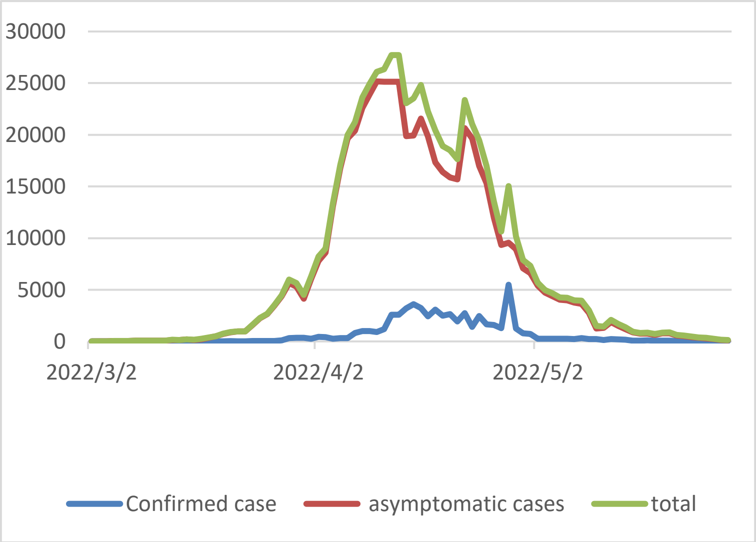

Figure 2. Daily new cases.

2.2. Overall data analysis of Weibo during the epidemic

The number of blogs posted in March, April, and May about this epidemic in Shanghai is shown in Figure 2. Overall, the daily number of blogs fluctuates wildly. The number of blogs posted on March 13 was twice that of the previous day, but it began to decrease in the following days. Since March 20, microblogs' number gradually rose in ups and downs. On April 2, it reached its peak, with 17694 influential daily blogs. Subsequently, the increase and decrease continued to alternate. After April 24, the overall trend was downward. By May 28, the general daily data was maintained, with a daily increase of only about 3000.

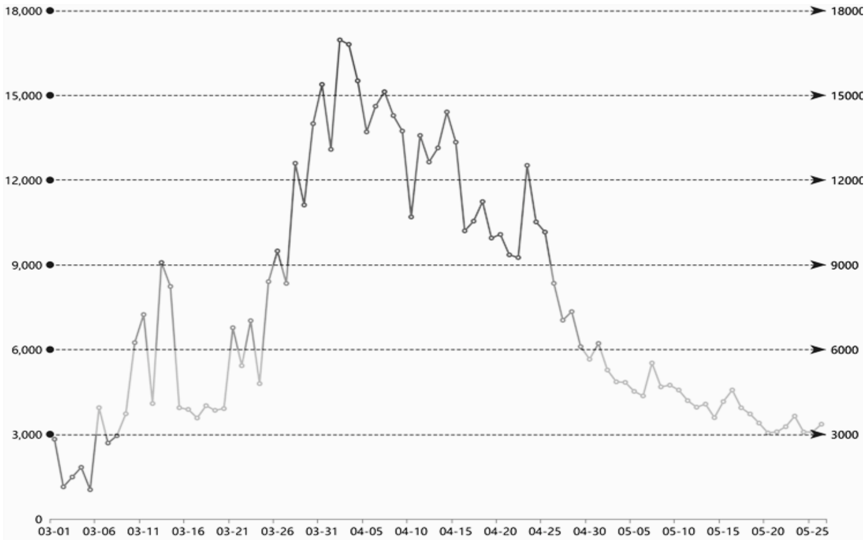

Figure 3. Number of microblogs posted every day.

3. Methods

The Bert model designed by Devlin J et al. has two main parts, Input Layer and Transformer Layer [14]. In the following sections, this paper first introduces Transformer, the essential part of the model. In the field of NLP, word embedding is the foundation of almost all research. Take word2vec (word to vector), proposed by Mikolov et al. in 2013, as an example [15]. It is mainly trained by the Continuous Bag-Of-Words Model (CBOW) and Continuous Skip-gram Model (Skip-gram). The CBOW algorithm's core idea is to cover a word in the sentence with a mask and then train the neural network to judge the masked word. Once the neural network can make the judgment accurately after a certain amount of training in the corpus training or reaches the established number of epochs, the word vector is calculated through the neural network. This is equivalent to encountering a word he cannot understand when reading an article and infers the meaning of the word from the context. As more and more complex natural language processing tasks emerge, researchers found that word2vec could not distinguish the conflicting semanteme of the same word, but in separate contexts. For instance, the word 'return' can mean 'come back', 'send back', and 'roll back', while word2vec merely outputs a one-word vector with a mixture of semantics. Therefore, a new method of word embedding - word embedding from the language model (ELMo) was proposed [16]. The innovation of this model is that it uses the bidirectional Long Short-term Memory (BiLSTM) based on the pre-trained word embedding model to encode the context of the word. The word embedding, which has been adjusted dynamically, considers the context's semantics. Each input word will go through three coding layers. The original word vector representation is inputted into the embedding layer, then goes through the corresponding lexical encoding in the forward LSTM layer, whose main storage is syntactic information, and finally goes through the corresponding lexical encoding in the reverse LSTM layer, which mainly includes semantic information.

In the model, each layer will be given a weight. The encoded results of each layer will be multiplied by their weights correspondingly, then they are integrated and summed to obtain the output of ELMo. Bert was launched immediately after Elmo. The fundamental change is that Bert replaces the original LSTM with a transformer encoder. The external structure of the transformer is composed of 6 encoders and six decoders stacked [17]. It receives the sequence data and then outputs the processed sequence data. The data processed by six encoders will be the input to six decoders for decoding.

In this model, every Encoder has the same structure, but the parameters are different from each other. It consists mainly of two parts -- Feedforward Neural Network and Multihead Attention Layer.

Decoder has a homologous interior structure and one more Masked Multihead Attention Layer than the Encoder to enable decoders to focus more on the critical parts of a sentence and ignore other less relevant words.

Taking Chinese as an example, this paper introduces the operation process of the Transformer. The sentence input is ‘The epidemic has come raging. It has caused significant economic losses. After processing, the original sentence is entered as every single character in the sentence. What the model needs to achieve is dealing with the input to understand that ‘it’ refers to an epidemic rather than something else.

When the model processes a word, the Self-attention layer calculates the degree of correlation between each of the two words in the sentence, in turn use the Softmax function to normalize the results to get the probability distribution of the degree. To clarify this process, this paper introduces relevant formulas to illustrate it. Set the input to the model to be X:

$$\mathrm{X} = (\mathrm{x}_0, \mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{N}) \tag{1}$$

And the output of the model to be Y:

$$\mathrm{Y} = (\mathrm{y}_0, \mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_\mathrm{N}) \tag{2}$$

The model creates linear mapping of each input, then multiplies it by the WQ matrix to attain the Query vector for each input. Similarly, it multiplies the linear mapping by the Wk and Wv matrix to get the Key vector and the Value vector for each input. These vectors are utilized to figure out the degree of correlation between each of the two words in the sentence thereupon. The words in the example sentence, 'It', 'has', are embedded into vectors x1, x2. The input X is multiplied by the corresponding weight matrix WQ/WK/WV to get the Q/K/V vector, then q1 and k1,k2,...,kT are correspondingly multiplied to get the Score of word correlation. After that, it divides Score by 8 (the square root of the K vector dimension) to get a new representation of attention. The result is normalized through the Softmax function to reduce the attention on the unimportant words while keeping the attention focused on the important characters solid. Then it multiplies the calculation result generated via the Softmax function by the value vector, and also combine the attention with semantic information to obtain the correlation vector z1 between 'It' and other characters in the sentence. The series of processes can be represented by a formula:

$$Z = Attention(Q, K, V) = softmax\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V. \tag{3}$$

Where d_k represents the dimension of the vector representing the key. While the Bert model adopts Multi-headed Attention, compared with self-attention, it provides multiple Representation Subspaces for the attention layer to expand the ability of the model to focus on different locations. Each header is assigned one query weight matrix, one key weight matrix and one value weight matrix (a total of eight headers are used). These weight matrices are randomly generated at the beginning of training, and the encoders or decoders’ vectors from the lower layer are projected to different Representation Subspaces through training. For each header, the corresponding attention degree vectors z_1, z_2, z_3...,z_7 is calculated through the formula. After the eight vectors are spliced and then multiplied by the matrix W_0, the model can obtain the final attention matrix Z. The above process can be expressed by formula:

$$Z = Concat(head0, head1, \dots, head7) \cdot W_0 \quad (4)$$

Where $head_i = Attention(Q_i, K_i, V_i)$

The Transformer used by Bert is modified based on the traditional model by adding residual connections. It concatenates each Encoder's and each Decoder's input with the results feedforward neural network or multi-head attention to deal with the problem of gradient vanishing as the network deepens. After that, the model update Z with a normalization function over the concatenation result, which aims to keep the output of the attention layer within specified bounds.

The series of processes can be represented by a formula:

$$Z \leftarrow LayerNorm\left(Concat\big(Z, Sublayer(X)\big)\right). \quad (5)$$

After the above procedure, the model inputs the normalization result into the neural network, and chooses Rectified Linear Unit (ReLU) to be the activation function.The procedure can be expressed by a formula:

$$Res(x) = max(0, xW_1 + b_1)W_2 + b_2 \quad (6)$$

n the formula, x represents the input and W1, W2, b1, b2 represent the parameters to be trained in the network.

Before Transformer, each character needs to be processed by the word embedding layer into a word vector provided to the Encoder. Unlike the Recurrent Neural Network, the Encoder only uses Multi-head Attention and feedforward neural network. Therefore, when the word vectors pass through it, it also needs to obtain the position information of the character to be encoded.

Positional Encoding uses sine function and cosine function to construct the value of each position, as shown in the following formula:

$$PE_{(position,2i)} = sin\left(\frac{pos}{10000^{\frac{2i}{d_{mod}}}}\right) \quad (7)$$

$$PE_{(postion,2i+1)} = cos\left(\frac{pos}{10000^{\frac{2i}{d_{mod}}}}\right) \quad (8)$$

4. Results and analysis

4.1. Data preprocessing and evaluation criterion

The data for this article comes from the Weibo web client. Sina Weibo is a social software platform based on user relations. Users can spread information and realize real-time interaction through text, video, pictures, and other methods. It is characterized by solid interaction and wide diffusion. This study uses Python's Scrapy, a distributed crawler framework, to obtain the Weibo text data of the Shanghai epidemic, an unconventional emergency, from Weibo. The selected time is from March 1, 2022, to May 29, 2022. The main content collected includes user name, Weibo content, publishing time, likes, comments, and forwarding, and a total of 652840 pieces of data were crawled. These data are divided into two categories by labels: P, indicating that the emotion of the text is positive and N, indicating that the emotion is negative. This paper screen out more than 80000 pieces of data to form a dataset, which contains 83254 comments. Examples of sentence emotion labels are cited in Table 1:

Table 1. Example of sentence emotion labels.

Text	Label
In this epidemic, I saw the hard work of hospital workers and paid tribute to them!	P
Because of this bloody epidemic, I haven't been out of the house for a long time. I feel like people are going to be paralyzed.	N

This paper segments the posts' texts using the Jieba tokenizer. Before segmentation, all words in the emotional dictionary are added to the Jieba tokenizer to prevent emotional words from being divided falsely, which will affect the classification results of the model. Afterword classification, illegal characters and stop words will be deleted from the output, which can significantly improve the model's accuracy. This paper selects three parameters as the evaluation criteria: Precision, Recall, and F1-measure. These have been used in the model evaluation of text classification and sentimental analysis tasks.

4.2. Baselines comparison

In order to fully reflect the capacity of the BERT model in respect of text classification and sentiment analysis, this paper selected some baseline models as contrasts for comparative experiments: CNN, a baseline model based on a feedforward neural network applied to public sentiment analysis. CNN+Att is a model which introduces an attention mechanism to features extracted from the neural network, giving each feature different weights to highlight more essential features and ignoring features with relatively low effect. It successfully enhances emotion classification ability: BiLSTM, an improved model using a two-layer LSTM layer. First, the input data is fed into the forward LSTM layer. Then the reverse time series is input into the reverse LSTM layer. Applying the LSTM model twice can improve the dependence of the model, thus improving the accuracy of the model. BiLSTM+CNN, a text classification model proposed by Yang et al., uses bidirectional LSTM to connect CNN, sets the length of the one-dimensional convolution kernel of CNN to 3, the number of convolution kernels to 64, and combines the maximum pooling layer with the softmax normalization function to form the classifier [18].

4.3. Experimental results

4.3.1. Parameter setting. This paper randomly divided the marked data sets into the training set and test set at a ratio of 5:5. This paper use the Chinese BERT-BASE model provided by the Huggingface official website for the pre-training and fine-tune it. Finally, a classifier is designed which is suitable for the data in this experiment. According to the introduction on hyperparameter selection by the original BERT paper, this paper set up parameters of 500 epochs, the learning rate of 5e-4, and the batch size of 16, as shown in Table 2 [19]. This paper uses the train set to train a two-category sentiment classifier with Softmax neural network layer.

Table 2. Parameter setting.

Parameters	Set Value/Function
Batch Size	16
Learning Rate	5e-4
Optimizer	AdamW
epoch	500
loss	Cross Entropy Loss

4.3.2. Sentiment classification results. This paper study the influence of the number of iterations on the classification results firstly. In terms of structure, it can be inferred that the number of iterations will significantly impact the model's performance. From Table 3, it can be clearly seen that the model's performance is optimal in the ninth iteration. When the number of iterations is less than nine, the model has not fully learned the features, so the performance cannot be optimal. When the number of iterations is more significant than nine, the model's performance gradually deteriorates due to overfitting.

Table 3. Model performance changing with number of epoch.

Epoch	Precision	Recall	F1	Epoch
3	95.2	88.9	91.9	3
5	95.4	88.2	91.7	5
7	95.5	88.5	91.9	7
8	94.9	91.4	93.1	8
9	94.5	92.0	93.2	9
10	91.2	92.1	91.6	10
11	92.3	93.5	92.8	11
13	92.9	90.5	91.7	13

Next, this paper study the influence of text attributes on the model's performance. It is found that the length of the text has a specific impact on the model's performance in the text classification task. Statistically, this paper selects the maximum text length, 599, and the average text length, 13, for research. Through experiments, the model is proved to be higher-performance when processing texts with average length, as shown in Table 4. This is because the model is fine-tuned to achieve Weibo's text classification. At the same time, Weibo data is primarily a short text. Once the text is too long, it will cause more time loss, thus reducing the model's performance.

Table 4. Model performance changing with sentence length.

Sentence Length	Precision	Recall	F1
13	93.5%	91.1%	92.3%
599	93.6%	82.0%	87.4%

Table 5 displays the results of BERT’s comparative tests against another baseline model. It seems clear that BERT performs better than other models in this public sentiment analysis task. This is due to BERT’s powerful performance as a pre-trained language model. Unlike traditional models, Word2vec, for instance, BERT can create word vectors flexibly and utilize them to complete some downstream text classification tasks.

Table 5. Performance comparison.

Model	Precision	Recall	F1	Model
CNN	74.2%	72.9%	73.5%	CNN
CNN+Att	78.3%	76.2%	77.2%	CNN+Att
BiLSTM	84.5%	82.9%	83.7%	BiLSTM
BiLSTM+CNN	89.5%	87.6%	88.5%	BiLSTM+CNN
BERT	92.1%	90.2%	91.1%	BERT

Finally, this paper takes negative emotions as an example to analyze the emotional changes online in the epidemic in Shanghai as a whole. It is apparent in Figure 4 that although the negative emotions of Internet users fluctuate, the overall proportion is high. From March 2 to March 6, there was a straight-line upward trend, and the proportion of negative emotions reached 60%. Most people expressed dissatisfaction, a large part of which was due to their poor understanding of the policy and lack of community logistics. Since then, there were ups and downs for more than a dozen days, with a decrease on March 14 and an increase on March 16. In the following period, the overall negative feelings declined in a weak trend. By May 26, they had dropped to about 36%. During the whole process, the peak of negative emotions of the public occurred in the early stage of the epidemic.

The first significant decline in negative feelings occurred around March 14. Combined with the epidemic situation, this is closely related to the short-term decline of the epidemic on March 13 after several days of continuous growth. Due to the beginning of precise prevention and control measures, most netizens changed their negative emotions, and the overall negative sentiment online decreased sharply. For example, a netizen posted a blog post:"# Shanghai epidemic # I believe the epidemic will get better soon. Shanghai should work hard!". Another netizen posted: "# Shanghai epidemic # I have never met such a good city. All requirements are met. I do not accept any negative news. I love Shanghai". However, the new data on the epidemic then rose again, and negative feelings begun to rebound.

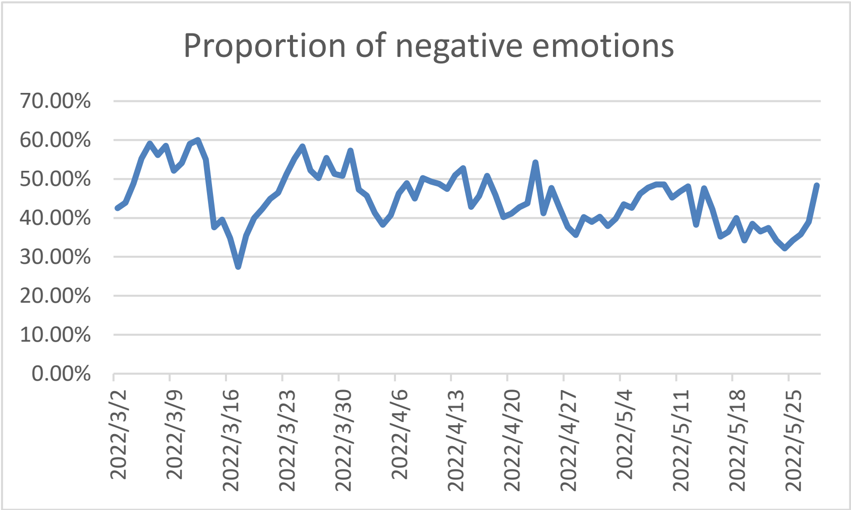

Figure 4. Proportion of negative emotions.

5. Conclusion

This paper uses the BERT model to analyze public sentiment during the epidemic. Taking the epidemic situation in Shanghai,2022, as an example, this paper crawls and analyzes the text, release time, number of likes, and other information of relevant blogs on Sina Weibo during this period and studies the basic development of network public sentiment. Then this paper uses the BERT model to systematically analyze the evolution process of network public sentiment in this epidemic to find the development trend of netizens' emotions when facing public health emergencies and explore the correlation and development law between network public sentiment and public events. Finally, some guidance can be summarized for dealing with similar public issues.

In addition, this paper describes the development history of word embedding technology, which leads to the BERT model's overall structure and operation process. At the same time, the whole process of using the fine-tuned BERT model to deal with the problem of public sentiment analysis is introduced, followed by the performance and classification results of this model in the sentiment classification of Weibo posts. Further, the performance of BERT is compared with other models, and the reason why this model is superior to other models is analyzed. BERT can also show strong performance when dealing with the sentiment classification of other texts.

This paper finds that because Weibo posts belong to Chinese language materials, some semantic information is ignored compared with English language materials. This is because BERT can only receive Chinese characters as input in this task. Therefore, the model can be optimized by designing an additional character recombine by imitating the structure of the Transformer to encode the input dynamically. After optimization, BERT will be competent for more complex sentiment classification tasks.

References

[1] Zhu Huayu,Sun Zhengxing,Zhang Fuyan. An Automatic Chinese Text Classification System Based on Vector Space Model [J]. Computer Engineering, (02):15-17+63.

[2] Wang Shumei,Dai Baocun,Wu Huizhong,Wang Fei. Text classification dictionary generation [J], Journal of Nanjing University of Science and Technology (Natural Science Edition),2002(05):517-521.
[3] Zhang Huanjiong, Li Yujian,Zhong Yixin. A New Method for Calculating Text Similarity, Law [J]. Computer Science,2002(07):92-93.
[4] KIM Y. Convolutional neural networks for sentence classification[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. New York: ACM,2014:1746-1751.2014.9.3.
[5] KALCHBRENNER N, GREFENSTETTE E, BLUNSOM P.A convolutional neural network for modelling sentences[C]/770-778.2015.12.10.
[6] Sun Mingmin.Chinese Text Classification Based on GRU Attention [J]. Modern Information Technology,2019,3(03):10-12.
[7] Liu Yue, Zhai Donghai, Ren Qingning News text classification based on attention CNLSTM network model [J/OL] Computer Engineering: 1-7 [2019-05-21] https://doi.org/10.19678/j.issn.1000-3428.0051312.2019.4.3.
[8] Liu Xinhui,Chen Wenshi, Zhou Ai,Chen Fei,Qu Wen,Lu Mingyu. Research on multi label text classification based on joint model [J/OL]. Computer Engineering and Application:1-10,[2019-07-18].http://kns.cnki.net/kcms/detail/11.2127.TP.20190624.1809.020.html.2019.6.25.
[9] Bo Tao, Li Xiaojun, Chen Su, Wang Yuting, Qi Guoliang Rapid seismic intensity assessment method based on social media data [J] Earthquake Engineering and Engineering Vibration, 2018, 38 (5): 206-215
[10] Zhao, Y. and Fan, B. (2018) Exploring Open Government Data Capacity of Government Agency: Based on the Resource-Based Theory. Gov-ernment Information Quarterly, 35, 1-12.
[11] Mergel, I., Kleibrink, A. and Sörvik, J. (2018) Open Data Outcomes: US Cities between Product and Process Innovation. Government Infor-mation Quarterly, 35, 622-632.
[12] Li Zongmin, Zhang Qi, Du Xinyu A study on the effect of rumor dispelling based on the interaction of rumor dispelling microblogs and the emotional tendency of popular comments -- Take the rumor dispelling microblogs related to the Xinguan epidemic as an example [J] Information magazine, 2020, 39 (11): 90-95+110
[13] Gao, J., Zheng, P., Jia, Y., Chen, H., Mao, Y., Chen, S., Dai, J., et al. (2020) Mental Health Problems and Social Media Exposure during COVID-19 Out-break. PLOS ONE, 15, e0231924.
[14] Devlin J ,Chang M W ,Lee K ,et al. BERT :Pre-tr aining of Deep Bidirectional Transformers for Language Und erstanding[J].2018.
[15] Tomas Mikolov ,Kai Chen ,Greg Corrado ,and Jeff rey Dean. Efficient Estimation of Word Representations in V ector Space.In Proceedings of Workshop at ICLR ,2013.
[16] Peters M E,Neumann M,Iyyer M,et al. Deep contextualized word representations[J]. 2018.
[17] Vaswani A ,Shazeer N ,Parmar N ,et al.Attention I s All You Need [J]. 2017.
[18] Yang Yan, Xu Bing, Yang Muyun, Zhao Jingjing. An emotion classification method based on the joint depth learning model [J]. Journal of Shandong University (Science Edition), 2017, 52 (09): 19-25
[19] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language Understanding", Proc. NAACL-HLT, pp. 4171-4186, 2019.