Emotional analysis model of college students based on big data public opinion

Shichao Luo

College of Mathematics and Physics, Leshan Normal University, Leshan, 614000, China

Abstract: With the development of Chinese social media, the occurrence of major events and social disputes can quickly produce a large number of text data, and academic and industrial circles pay more and more attention to emotional analysis of public opinion events. The contents involved in university online public opinion are often the real-time reflection of students' various ideological trends. Because the psychology and physiology of college students are not mature enough, in the process of online public opinion events, extreme speech and behavior are often produced due to the stimulation of various external factors, which has a negative impact on the harmony and stability of universities. In this paper, an emotional analysis model of college students based on big data public opinion is designed. Through the technology of web crawler, Word2vec model is adopted to optimize the emotional dictionary, CNN-LSTM (Convolutional Neural Network-Long short-term memory) carries out text emotion classification module, and the emotional analysis of user comments is carried out. According to the analysis of students' public opinion, the effective prediction of college students' excessive behavior is realized, which significantly improves the quality of students' mental health.

1. Introduction

When people express their views on something or things online, they often express their positions, attitudes and emotions through some sentences. With the development of Chinese social media, the occurrence of major events and social disputes can quickly produce a huge amount of text data, and academic and industrial circles pay more and more attention to emotional analysis of public opinion events [1]. Sentiment analysis, as an important module of public opinion analysis, is mainly responsible for extracting emotional text information from a large number of original text data and using the obtained text emotions for the next text classification [2].

Compared with newspapers and other media, the Internet has the advantages of digital products, networked circulation and convenient distribution. At the same time, there are also some problems such as difficult information identification, numerous network rumors, and difficult semantic analysis. Therefore, it is of great significance to strengthen the analysis of network public opinion in order to create a good network space and grasp the correct direction of public opinion [3-4]. This paper puts forward a model of emotional analysis of college students based on Internet big data and social platform resources. Through information technology and big data processing technology, the relevant data obtained in the Internet are analyzed, and through the interpretation of emotions, a complete mental health evaluation system is constructed, so as to effectively predict the excessive behavior of college students.

2. Emotional analysis of big data public opinion

The analysis of public opinion and emotion is mainly based on the attitudes and emotions expressed in public opinion, and analyzes people's attitudes towards some current social phenomena and social events. At present, online media has become the main position for people to express their views. Through the articles published on the media platform and the comments published by netizens, we can analyze their emotional attitudes towards social events and further analyze their views through attitudes. Emotional analysis is mainly divided into methods based on emotional dictionaries and methods based on machine learning. The method based on emotional dictionary is

more traditional, which is mainly based on the emotional dictionary summarized by experts in the psychological field, and accumulates emotional scores by identifying emotional words in the text. At present, there is still a lack of public opinion and emotional analysis for online platforms such as portals [5]. Based on this, this paper makes a comparative study of public opinion and emotion on multimedia platforms.

The contents involved in university online public opinion are often the real-time reflection of students' various ideological trends. Because the psychology and physiology of college students are not mature enough, in the process of online public opinion events, extreme speech and behavior are often produced due to the stimulation of various external factors, which has a negative impact on the harmony and stability of universities. Public opinion analysis of higher education is the expression of public's emotions, attitudes and behavioral tendencies on educational issues, the public's demands for educational rights and interests, and it contains the complex mentality of various social interest groups on educational issues.

Emotion analysis is an extremely critical module in public opinion analysis. If emotion and the underlying reasons can be extracted in the emotion analysis stage, the precision of public opinion analysis will be greatly improved [6]. Therefore, we add the emotional reason extraction based on causal analysis to the public opinion analysis system, so that the whole public opinion analysis system can get rid of the problem of manual labeling, improve the efficiency of the public opinion analysis system and reduce the economic expenses.

3. Construction of students' emotional analysis model

3.1. Overall design process

Text emotion clustering based on emotion types can not only fully display educational public opinion emotion at the macro level, but also clearly show the characteristics of data at the micro level, which can effectively help users to analyze their emotions. In current society, there are a large number of users who need to pay attention to education, such as parents and teachers, educators and online opinion leaders, who want to know the educational information published in various online media. Employees in the education industry also need information based on various public opinion data to help adjust their institutional business, that is, they want to know the current opinions of parents on various educational activities and their online communities and other data, so as to adjust the advertising and publicity content of educational services.

As an important part of netizens, college students can easily obtain all kinds of information by means of the network platform, and can express their attitudes and opinions related to their own interests or on various social public hot events relatively freely, which has become the main generating force and influencing object of online public opinion [7-8]. In the field of Chinese social media, the text has some characteristics, such as incomplete grammatical structure and too many network neologisms, which are not conducive to the algorithm to accurately complete the text classification task. In the face of massive data, the algorithm is inefficient, which brings great challenges to public opinion supervision. Ensemble learning algorithm can solve the problem of low precision in text classification task. Spark is programmed by Scala, and supports Python and Java programming languages to complete development tasks through API access. It inherits the advantages of high fault tolerance and strong scalability of MapReduce.

Semantic analysis of texts, pictures and multimedia data in online Internet resources through information technology can be used to understand people's attitudes, opinions and emotional expressions [9]. The emotional analysis model of college students based on big data public opinion proposed in this paper is a big data ontology collected from social networking sites through mining technology, which carries out semantic analysis on semi-structured and unstructured data with data types such as text, pictures and videos, and deeply digs the fluctuation of students' mental health index under emotional changes, so as to effectively predict the possible excessive behavior tendency. The emotional analysis model of college students based on big data public opinion is shown in Figure 1.
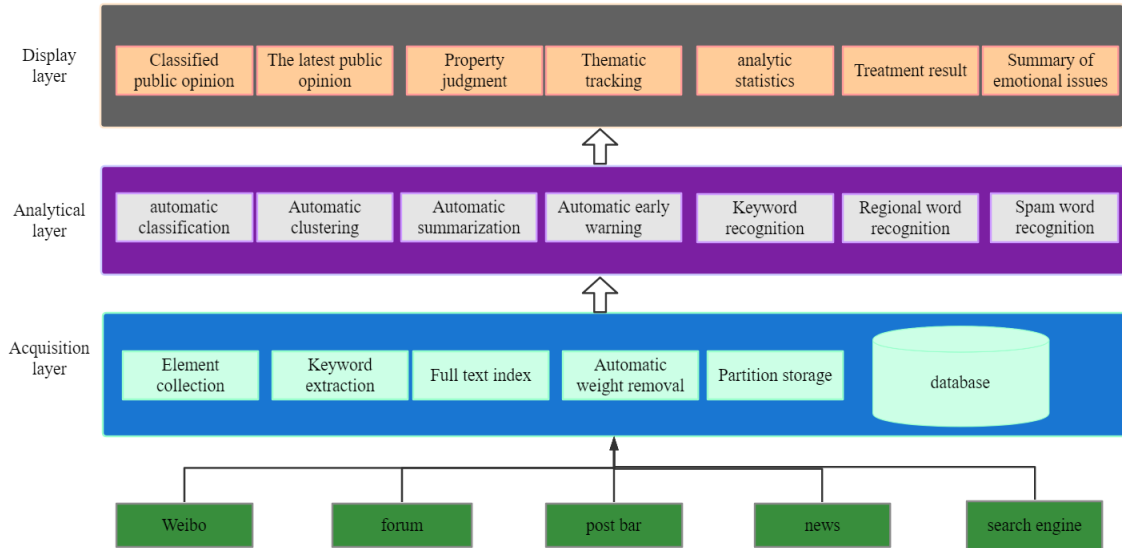
Figure 1 Emotional analysis model of college students based on big data public opinion

Using the distributed framework of big data, a distributed public opinion data collection mechanism is established to improve the efficiency of data acquisition. At the same time, unstructured data can be directly captured, which greatly saves the time of relational database conversion processing. Firstly, the public opinion information to be processed is divided into a plurality of data blocks distributed on different data nodes; Then, corresponding to each data block, the data processing job is divided into a plurality of calculation tasks to be executed; Finally, the processing results are summarized in the central node. This method ensures the reliability and expansibility of execution.

3.2. Analysis of students' emotion

When users express their opinions, the posts of forum information sources, especially the replies, are varied in sentence patterns and modification techniques, and sometimes they are mixed with irony and satire, which the computer can not automatically and accurately understand. It can make neural network pay more attention to the part of the text that contains more emotional information. Finally, we analyze the processed text information based on causality, extract individual emotions and reasons through multi-task learning, then carry out emotion-reason pairing and filtering, extract keywords to get emotional grades, and finally make emotional grade line charts and word cloud charts for visual display of data.

In order to classify Chinese emotions, emotions are divided into seven categories, and under the guidance of domain experts, the seven categories of emotions are sorted from positive to negative: joy, kindness, shock, sadness, fear, evil and anger. In addition, in order to improve the precision of emotion recognition, a degree adverb dictionary is added to calculate the weight of emotion. In order to analyze the emotional tendency of public opinion text, this paper uses the method of accumulating emotional words to identify the emotion of the document.

Based on the emotional dictionary, this paper combines adverbs and negative words to modify emotional words, and introduces the modification distance to judge the influence of modified words on emotion, which improves the precision of emotional recognition [10]. We express the emotional score of a sentence as a 7-dimensional vector according to 7 emotional types, such as the formula:

$$E(S) = \left\{ e_i = \sum_{j=1}^{n} score * weight \Big| index_{\min} \leq i \leq index_{\max}, i \in N^+ \right\} \tag{1}$$

The 7-dimensional emotional score vector of sentence S is expressed as $E(S)$. In this paper, $index_{\min} = 1, ndex_{\max} = 7$ means to calculate all seven emotional categories. i represents the index of

emotion category to be calculated, and e_i represents the intensity of a specific emotion category. j represents the number of emotional words in the sentence.

By accumulating the emotional intensity of all sentences in the document, the emotional vector $E(F)$ of the document is obtained. As shown in Formula (2):

$$E(F) = \sum E(S_n) \quad (2)$$

Among them, $E(S_n)$ is expressed as the 7-dimensional emotion vector of the n sentence in the document.

In order to improve the precision of sentiment analysis, we further adopt Word2vec model to optimize the sentiment dictionary and analyze the sentiment of user comments. Word2vec algorithm is essentially a simple two-layer neural network, which mainly includes skip-gram model and CBOW model. The algorithm flow of skip-gram is to input a one-hot encoder word vector of a specific word into the neural network.

The main idea of CBOW model is the opposite: input the one-hot encoder word vector of the context word corresponding to a specific word, and get the word vector of the specific word through training. The essential mathematical expression of the model is:

$$P\left(w_{(t)} \middle| w_{(t-k)}, w_{(t-k+1)}, \cdots, w_{(t-k-1)}, w_{(t+k)}\right) \quad (3)$$

Among them, $w_{(t)}$ is the t-th word in the text, and k is half of the total contextual words of the word $w_{(t)}$.

For the study of Chinese text emotion classification, CNN (Convolutional Neural Network)is better at short text classification, while LSTM (Long short-term memory) is better at long text classification. So this chapter constructs CNN-LSTM model based on this. The main body of the model is mainly divided into four parts, the first is the input part of the model, the second is the extraction part of emotional features, the second is the fusion part of the features extracted by the two models, and the last is the output part of the model.

Xavier is the method used in initialization in this paper. Assuming that $X_{i:i+h-1}$ represents the data from the i-th row to the $i+1$-th row in the matrix, and the convolution kernel is W, the processing process of the convolution kernel with height h is shown in Formula (4).

$$c_i = f\left(W \cdot X_{i:i+h-1} + b\right) \quad (4)$$

The introduction of Attention mechanism will make some information with high influence give more attention in order to ignore the influence of some unimportant information. Calculate the similarity between the randomly initialized text input under the Attention mechanism and the words output from the previous LSTM layer, as shown in Formula (5).

$$simi\left(\overline{g}, g_i\right) = \tanh\left(W \cdot \overline{g} + U \cdot h_i + b\right) \quad (5)$$

Finally, the selection of optimization function. Adagrad function is used in this paper. When the training number is t, the method of updating each weight factor of the model is as shown in Formula (6).

$$\Theta_t = \Theta_{t-1} - \frac{\alpha}{\sqrt{\sum_{i=1}^{t} g_i}} \quad (6)$$

Θ is the set of all parameters to be trained, α is the learning rate, and g is the gradient of the t epoch.

Initialize the extended dictionary as the original emotion dictionary. 50% preprocessed text data

is used as training data set, which is input into Word2vec model to get the word vector of text data. Then, the sum of the intensity of different types of emotional words in the emotional word subset is accumulated, and the emotional type with the greatest intensity is the emotional type of the word, and the emotional intensity of the word is the arithmetic average of the emotional intensity of all the words of this type in the emotional word subset.

4. Experimental analysis

We used our own web crawler program to publish the popular comments made by other users during the epidemic prevention and control in COVID-19 on Weibo platform, and initially collected about 20,000 user comments. The crawled initial data has some problems, such as redundancy and nonstandard format. After simple preprocessing, about 15,000 pieces of standardized data are obtained after deleting relevant nonstandard data. We use Python language to preprocess the text.

For the topic of "COVID-19 epidemic", the review data are classified into three categories: positive emotions, negative emotions and neutral emotions by using the Rost tool, and the results are drawn into a pie chart. As shown in Figure 3:
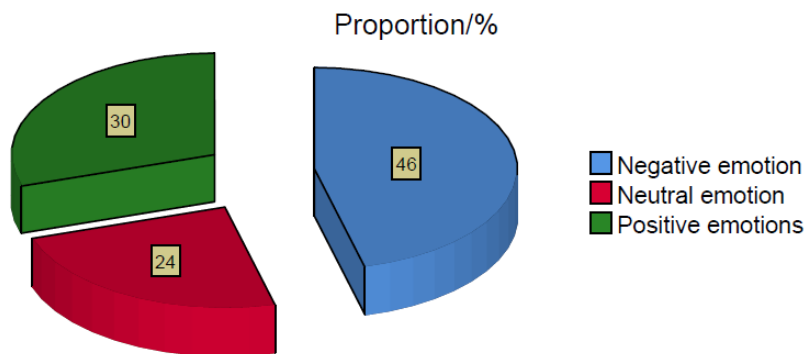

Figure 3 Topic users comment on emotional analysis results

As can be seen from Figure 3, in the topical events of "COVID-19 Epidemic", negative emotions account for 46%, neutral emotions and positive emotions account for 24% and 30% respectively. By observing the comment data of negative emotions and analyzing the historical information of users in Weibo, it is found that these comment data show negative emotions. Most users in Weibo are college students whose dormitories have been requisitioned.

In order to further reflect the emotion classification performance of CNN-LSTM model proposed in this paper, the precision rate, recall rate and F1 value are used as evaluation indexes. The experimental results are shown in Table 1 and Figure 4.

It can be seen that the effect of using CNN-LSTM as a feature extraction model is the best, and the F1 value can reach 0.897, and the precision and recall rate are the best. Through experiments, the model proposed in this paper is better than other models in precision, recall, F1 value and overall classification precision.

In the current Internet era with social media as the mainstream, university network public opinion has become an important part of social network public opinion. University online public opinion has the characteristics of many sensitive events, rapid spread, high complexity and wide influence, which increases the difficulty of online public opinion management. Therefore, universities need to increase technical input, actively and scientifically use big data technology, establish a database of public opinion events, build a school-wide student management data platform, and establish data models such as "public opinion quantitative index system", "opinion tendency analysis model" and "evolution analysis model" to find useful information for public opinion prediction and judgment.

Table 1 Experimental result

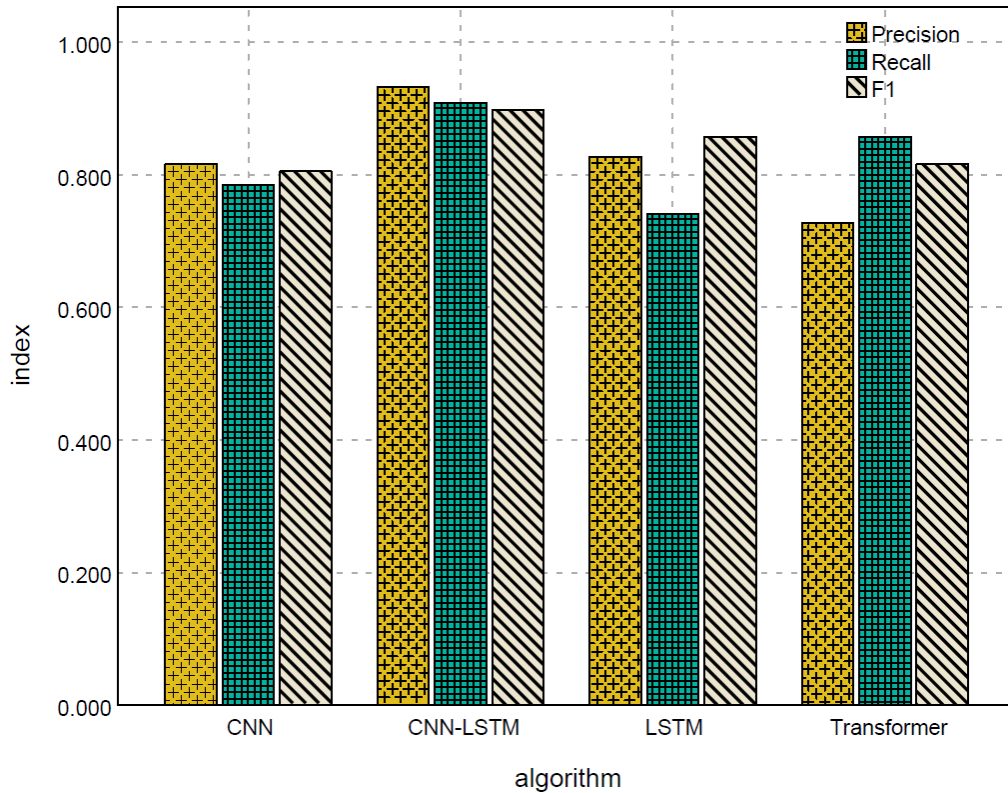| algorithm	Precision	Recall	F1
Transformer	0.727	0.857	0.816
CNN	0.816	0.785	0.805
LSTM	0.827	0.741	0.857
CNN-LSTM	0.932	0.908	0.897

Figure 4 Statistics of experimental results

5. Conclusions

Sentiment analysis, as an important module of public opinion analysis, is mainly responsible for extracting emotional text information from a large number of original text data, and using the obtained text emotions for the next text classification. University online public opinion has the characteristics of many sensitive events, rapid spread, high complexity and wide influence, which increases the difficulty of online public opinion management. Therefore, this paper puts forward a model for analyzing college students' emotions based on Internet big data and social platform resources. Through information technology and big data processing technology, the relevant data obtained in the Internet are analyzed, and through the interpretation of emotions, a complete mental health evaluation system is constructed, so as to effectively predict the excessive behavior of college students.

References

[1] He Yue, Zhu Tingting. Research on smog public opinion based on Weibo's emotional analysis and social network analysis [J]. Information Science, 2018, 36(7):7.

[2] Zhang Chen, Ma Xiangyuan, Zhou Yang, et al. Analysis on the evolution of public opinion of COVID-19 epidemic based on users' emotional changes [J]. Journal of Geo-information Science, 2021, 23(2):10.

[3] Jiang Zhiyi, Ma Wangrong, Zou Kai, et al. Study on the emotional evolution characteristics of online public opinion based on the analysis of emotional tendency [J]. Modern Intelligence, 2018, 38(4):8.

[4] Qiu Zeguo, He Baiyan. Network public opinion clustering and emotional evolution analysis based on PCA-Spectral-LDA: A study of Weibo text mining [J]. System Science and Mathematics, 2021, 41(10):13.

[5] Andy Hu, Shu Xinggang. Analysis of public opinion emotional tendency of community group buying media based on knowledge map [J]. Financial Theory and Practice, 2021, 042(002):119-124.

[6] Tsaih R H, Kuo B S, Lin T H, et al. The use of big data analytics to predict the foreign exchange rate based on public media: A machine-learning experiment[J]. It Professional, 2018, 20(2):34-41.

[7] Kong Jingyuan. Research on the influence of the response of public opinion parties on the emotion of netizens [J]. Library and Information Work, 2020, 64(18):89-96.

[8] Jia F, Chen C C. Emotional characteristics and time series analysis of Internet public opinion participants based on emotional feature words:[J]. International Journal of Advanced Robotic Systems, 2020, 17(1):105-108.

[9] Cla B, Jb A, Lei Z B, et al. Opinion community detection and opinion leader detection based on text information and network topology in cloud environment - ScienceDirect[J]. Information Sciences, 2019, 504(C):61-83.

[10] Ghanem B, Rosso P, Rangel F. An Emotional Analysis of False Information in Social Media and News Articles[J]. ACM Transactions on Internet Technology, 2020, 20(2):1-18.