Research on the Monitoring and Analysis Method of College Students' Public Opinion Based on Bert-CNN Model

Changjin Qian[1], Pingping Fang[2, *]

[1]School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu 233000, China;

[2]School of Literature, Anhui University of Finance and Economics, Bengbu 233000, China.

* Corresponding Author

Abstract

With the rapid development of Internet technology and the strengthening of media diversification trend, the Internet has become the main carrier of public opinion information. How to effectively and accurately obtain the emotional tendency from the network public opinion text, so as to accurately control the direction of public opinion has become an important topic in the Internet era.College students, as a group with a high rate of Internet activity, are the main occurrence and influence objects of public opinion. How to effectively manage and guide online public opinion in colleges and universities has become a necessary condition to maintain the safety and stability of colleges and universities and the healthy development of teachers and students. It is of great practical significance to monitor and analyze the network public opinion of colleges and universities in time.Based on this, this research is committed to building a Bert-CNN based college student public opinion monitoring and analysis model. Taking the postgraduate examination of Anhui University of Finance and Economics under the background of the epidemic as an example, the empirical analysis is carried out.Firstly, the topic text data on each mainstream network platform was crawled by python, and the word vector of the text data was generated by Bert and imported into CNN for feature extraction. In addition, Textrank algorithm was used to classify the public opinion text into three categories and generate the word cloud maps of different emotional categories. Secondly, the Louvain-Kmeans clustering model is used to conduct spatial clustering on the results of public opinion analysis, further analyze the commonalities and differences of regional public opinion on a larger scale, and finally obtain the emotional characteristics, regional correlation degree and focus of attention of relevant public opinion. Provide favorable data support and guarantee for managers to control public opinion guidance and the formulation and implementation of emergency handling measures.

Keywords

Bert-CNN; Online public opinion; Postgraduate examination; Current situation of public opinion.

1. Introduction

With the rapid development of computer network technology, major Internet platforms have become the main places for the generation and dissemination of public opinion in the information age. In recent years, the trend of media diversification has been increasing, and the Internet has become the main carrier of public opinion information. How to effectively, accurately and quickly obtain emotional tendencies from online public opinion text information, so as to more accurately and quickly control the direction of public opinion, has become an

important issue in the Internet era. Public opinion is a "barometer" of people's livelihood and reflects changes in social sentiment. (Yang Wenwei, 2015) The analysis of social public opinion can help relevant departments to accurately study and judge the development trend of public opinion communication, timely discover and deal with negative public opinion, identify the emotional attitude and opinion tendency hidden by netizens in network information, and better judge the trend of public opinion and cope with public opinion crisis. (Jiang Gang, 2017)

The advent of the era of big data makes the situation of Internet public opinion increasingly complex, and the main body of Internet use is diversified. Among them, students account for 21% of China's total Internet users. College students, as a group with high network utilization and activity, are the main generating and influencing objects of network public opinion, and college network public opinion often has the phenomenon of group polarization and spillover effect due to its distinct main characteristics. In recent years, there have been frequent online public opinion incidents in colleges and universities, and negative crises caused by campus safety accidents, teachers' ethics, and recruitment examinations have emerged one after another, which have had an impact on the harmony and stability of colleges and universities and society. Reasonable and effective guidance and management of online public opinion in colleges and universities can maintain the safety and stability of colleges and universities, promote the healthy development of teachers and students, and improve the ability of government and relevant departments in colleges and universities to cope with online public opinion and the level of governance, which is of great practical significance. (Zhu Xiaoxia, 2019)

However, the current domestic research on online public opinion in colleges and universities mainly focuses on the characteristics of public opinion, communication rules, coping strategies and guidance mechanisms, and lacks a complete and effective public opinion monitoring and analysis mechanism in the face of the frequent outbreak of public opinion events in colleges and universities and the aggravation of harm. Based on this, this study will conduct sentiment analysis on the extracted public opinion text based on the BERT-CNN model. Compared with the traditional language understanding model, we combine the high language understanding accuracy of BERT and the strong network level deep fitting ability of CNN, which can improve the prediction efficiency, accuracy and comprehensiveness of the model compared with the traditional method. It provides a general framework for deep learning model in the field of public opinion analysis of college students, and has important practical significance in effectively and accurately guiding the health direction of universities and public opinion.

2. Literature review

At present, the research on public opinion analysis in China is still in the preliminary stage, most of the research results belong to the applied research category, and the research on public opinion theory is still insufficient. At present, the focus of public opinion research in China is mainly focused on the basic theory and definition of public opinion. Public opinion collection, analysis mechanism and method; Psychological pressure and social public opinion transformation; In terms of the correlation between mass emergencies and public opinion, as well as between national decision-making and public opinion mechanism, and due to the relatively short development history of the Internet, China's research on online public opinion can be said to have just started.

Foreign research is more mature than domestic research. Since the late 19th century, public opinion research has attracted the attention and research of sociologists. Up to now, public opinion surveys are quite common in Western countries. Foreign research on public opinion mainly involves: participatory democracy, government decision-making and public opinion surveys, policy making, supervision by public opinion, etc.

Public opinion analysis is a new research field of social science, and its main research methods can be divided into two categories as a whole:

One is the big data type. From this perspective, public opinion research can be roughly divided into the traditional paradigm of social public opinion analysis and the network big data based public opinion analysis. The analysis of social public opinion in the traditional paradigm is more inclined to study the influencing factors of social public opinion, focusing on the formation and evolution mechanism of journalism, communication and sociology, and putting forward policy suggestions on this basis. For example, MacLennan (2013) sampled 2,337 residents aged 18 and above in different local communities about their views on alcohol and local government alcohol policies, and Alan (2016) et al. studied the relationship between people's dissatisfaction with leaders and the number of terrorist attacks in the Middle East and other places. After the rise of network media, researchers gradually shifted their attention to Internet social platforms. With the deepening of the research, researchers found that social media platforms such as Twitter, Facebook and Weibo have risen rapidly, so the research on network public opinion based on big data technology has attracted wide attention from experts and scholars. Ceron(2020) understands citizens' political preferences through sentiment analysis of Twitter text data, and proves that there is a significant correlation between social media and traditional public survey results. Kang Wei (2021) used big data technology to collect public opinion information from major news media and social media in "1.16 school bus accident", generated network relationship diagram and put forward constructive guidance opinions.

Second, it is combined with the type of mathematical model. With the deepening of the research, the combination of big data analysis and processing technology on the basis of inheriting the research paradigm of network public opinion analysis has gradually become the mainstream research method. Mehta,Signorini(2019) et al. used Twitter's official API to track the development of H1N1 influenza. Xiao S(2021) et al. built a web crawler based on multi-strategy access mechanism, breaking through the limitation of the original method of acquiring microblog text in the number of grabs. Some researchers cleaned the obtained online public opinion text and combined it with domain dictionary and similarity calculation to evaluate the Internet hot spots. Xia Huosong outlined the sentiment analysis methods used in online public opinion research, such as support vector machine (2022) and naive Bayes algorithm (2014). Alessio(2021) collected Twitter text data with H1N1 as the keyword and used the open-source libSVM for text classification. Ren Peng et al. (2022) applied SVM model to the study of public opinion text classification, and the corpus came from crawled microblog posts. Zhang Chen(2022) applied the artificial neural network technology based on big data to the risk control work in the financial field.

3. Relevant technologies and theories of applying text deep learning to public opinion monitoring and analysis

The existing methods of public opinion analysis mechanism in colleges and universities have a good performance in public opinion analysis and prediction, but there are still some shortcomings. For example, some public opinion indicators are difficult to evaluate and are highly subjective; Lack of public opinion indicator screening steps; There is a lack of empirical research and no actual cases to verify the rationality and effectiveness of predictive analysis methods.

Based on this, in order to enrich and improve the existing research, this study combines the characteristics and development and evolution rules of college students' public opinion, and on the basis of comprehensiveness and science, builds a BERT-CNN model and conducts emotional text analysis by combining TextRank algorithm and K-means clustering model.

Firstly, the existing literature studies on public opinion monitoring and analysis using BERT, neural network model and other methods will be summarized and sorted out. Based on the current research architecture, innovation will be made to determine the research direction and method adopted in this paper. Public opinion monitoring and analysis is a very necessary hot research issue in the Internet era. When deep learning is used for public opinion monitoring and analysis, there are two stages of work: first, feature extraction of relevant public opinion topics and content, and then emotion analysis identification and induction of the extracted features after learning the network model. At present, deep learning and neural networks are used to monitor and analyze public opinion data, mainly including BERT model, support vector machine model, BP neural network, LSTM, RNN and other network structures.

4. Data acquisition, data preprocessing and basic architecture introduction of Bert model

When conducting sentiment analysis of public opinion texts, it is necessary to collect relevant corpus resources, conduct pre-processing and text vectorization on these corpus resources, and finally use these vectorized texts to complete the training and testing of the model. The data used in this study is from November 2021 to January 2023. As data sources, the five major social media platforms, namely Weibo platform, wechat public account, Baidupost Bar, Zhihu and Qzone, are used to compile multi-threaded crawler tools using Python language and Scrapy framework to crawl text data from the above platforms. The results are stored in MySQL database for real-time public opinion detection. The pre-processed corpus is processed by the Jieba tool in Python for word segmentation.
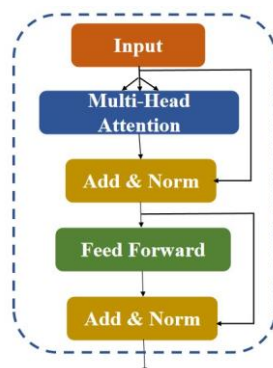

Figure 1 Structure diagram of Encoder

BERT uses the Encoder part of Transformer as its basic unit, and residuals are used to enable its network to reach a very deep degree, and it can more thoroughly capture the bidirectional relationship in the statement for bidirectional coding. The specific structure diagram is shown above.

5. A BERT-CNN student public opinion analysis model is proposed.

Based on the inductive analysis and summary of relevant research results at the present stage in the first two parts, a Bert-CNN model for public opinion analysis is constructed and trained for the current problems. BERT model, as the best text analysis model in the field of NLP, can extract the emotional features of public opinion text more accurately. First, we use the BERT model pre-trained on the Chinese corpus to train and generate word vectors. Generally speaking, BERT-large has more parameters than other models, so it is better to use, so we choose it as our pre-training model. The next part is feature extraction, which involves two

MaxPooling layers and a CNN model. The generated word vector is used as the input of the CNN model and MaxPooling layer, and a one-dimensional vector v1 output by MaxPooling layer is added to the CNN model. Splice with the one-dimensional vector v2 output by the first MaxPooling layer. The spliced vector is used as the input of the fully connected layer, and then the classification result is output. The MaxPooling layer combines the original features, maximizes the features and combines them with CNN, and retains the original feature information of the word vector to ensure more accurate results. Finally, the unsupervised TextRank algorithm is used to extract theme words and carry out emotion state and intensity analysis. Moreover, Louvain-Kmeans clustering is used to carry out spatial cluster analysis on the obtained emotion state data, and finally the emotional characteristics, regional correlation degree and focus of public opinion are obtained.
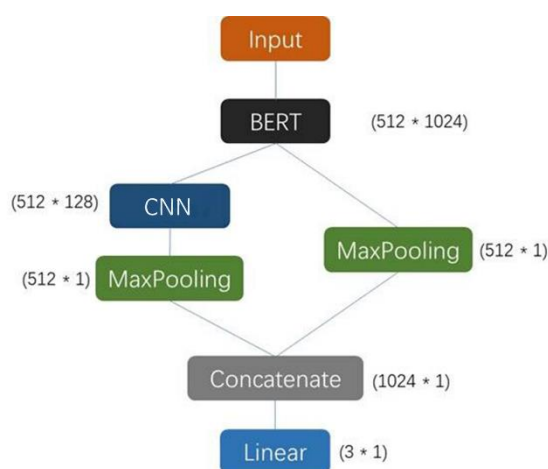

Figure 2: Model structure diagram

6. Empirical research on the monitoring and analysis of public opinion in university graduate examination based on the epidemic background

In this part, the information and data related to the postgraduate examination of Anhui University of Finance and Economics during the epidemic period are used for empirical research. Using Weibo and other well-known media platforms as data sources, Python was used to crawl the text data involving Anhui University of Finance and Economics, Ancai, postgraduate entrance examination and other key topics, web news, short videos and related comments from the above network platforms. The data is pre-processed and input into Bert model for vectorization of text. The correlation degree of words in sentences is fully considered. Vectorization results are input into CNN model for learning. TextRank algorithm is used to extract theme words and complete the text classification task (positive, neutral and negative) of public opinion analysis. And generate word cloud map according to keyword weight. By analyzing the information proportion of the word cloud map, the public opinion of different emotional types is identified. According to Louvain algorithm, the network space between regions is constructed for community discovery. By Kmeans, the communities are further grouped into a specific number of regional categories by integrating regional emotion and spatial attributes, and the commonalities and differences of regional public opinions on a larger scale are analyzed. Finally, according to the type, degree and spatial characteristics of public opinion, we can conclude that after the exam, public opinion is mainly reflected in the difficulty of the exam, the exam questions and the postgraduate life of Anhui University of Finance and

Economics; In terms of spatial characteristics, most of them are concentrated in the north-central part of Anhui Province and the south-western part of Jiangsu Province.

7. Conclusions and recommendations for improvement.

This paper summarizes the previous models of public opinion analysis, builds a BERT-CNN public opinion monitoring and analysis model, and proposes the problems that deep learning methods may face in the process of public opinion monitoring and analysis, as well as the possible improvement and optimization paths of the model structure. By analyzing the emotional attitude and opinion tendency of netizens, we can assist universities to formulate control strategies in a timely manner, guide the direction of public opinion, and avoid the occurrence of bad public opinion. This provides theoretical reference and method support for maintaining the stability of the network environment of universities, and also provides new ideas for the design and implementation of the public opinion monitoring and analysis system.

Acknowledgements

Reference

[1]   Chen Z,Xinmiao Z,Jiaqing W. [J]. Mobile Information Systems,2022,2022.

[2]   Wan Liwen. Enterprise public opinion based on the SVM classification and early warning research [J]. Modern commercial and trade industry, 2022 lancet (19) : 65-66. The DOI: 10.19311 / j.carol carroll nki. 1672-3198.2022.19.030.

[3]   Andrea C,Luigi C,Wiebke D. [J]. German Politics,202,31(3).

[4]   Shao Deqi, Feng Chao, Wang Liping. Characteristics and governance of Network public opinion from the perspective of new media [J]. China Media Science and Technology,2022(06):7-9.DOI:10.19483/j.cnki.11-4653/n.2022.06.001.

[5]   REN Peng, Li Wenjie, Shu Yujie et al. Research on news text classification based on BERT word embedding and bidirectional cyclic convolutional neural network [J]. Information recording materials, 2022, 23 (6) : 20-23, DOI: 10.16009 / j.carol carroll nki cn13-1295 / tq. 2022.06.073.

[6]   Alessio B,Alessandro B,Pietro D, et al. [J]. IEEE ACCESS,2021,9.

[7]   Ilha M C,Barichello M G,De R C O, et al. [J]. CoDAS, 2019,32(4).

[8]   Zhu Xiaoxia, Meng Jianfang, Song Jiaxin. Social public opinion in public emergency management choice - based on the theory of explosion seepage [J]. Journal of intelligence science, 2019, 5 (3) : 48 + 43-80 DOI: 10.13833 / j.i SSN. 1007-7634.2019.03.008.

[9]   Challenges and tasks of educational examination Informatization [J]. Chinese Journal of Examinations,2017(06):1-5.DOI:10.19360/j.cnki.11-3303/g4.2017.06.001.

[10] Burkitt-Gray A ,Burkitt-Gray A . [J]. Global Telecoms Business,2016.

[11] Xia Huosong, screening spring. Literature review on public opinion analysis and decision support in Big Data environment [J]. Journal of Information,2015,34(02):1-6+21.

[12] Xu J,Ma B. [J]. Applied Mechanics and Materials,2014,3009(519-520).

[13] Kang Wei. Social Network structure measurement and analysis of public opinion communication in emergencies -- Based on the empirical study of the "November 16 school bus Accident" [J]. Heilongjiang Education (Theory and Practice),2014(03):59-65.

[14] Brett M,Kypros K,Robin R, et al. [J]. Addiction (Abingdon, England),2013,108(5).

[15] Alessio S,Maria A S,M P P. [J]. PloS one,2011,6(5).