Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 2
CII 2023

Research on the Effect of Information Transmission based on Text Mining

Yuting Sun

School of science, Lanzhou University of Technology, 730000, China

19119389963@163.com

Abstract. Sangstan proposed that, in the process of information dissemination, people only focus on the information field that makes them feel comfortable due to the influence of the public's own information needs and interest guidance. As time passes, they will shackle themselves to the cocoon like "cocoon room", which Sangstan called "information cocoon room". As an emerging open social media, microblog provides users with the convenience of information sharing and dissemination, and also creates a personalized social space. Based on this, this paper conducts text mining on topics in microblog through text mining technology, aiming to study the propagation law of public opinion. For Task 1, this paper uses the crawler technology to access the microblog search end api interface, and the code uses the python request package, so that the access interface can be determined and the data can be crawled. After that, this paper uses the jieba library to carry out regular expression and Chinese word segmentation, and constructs a word cloud map for statistical analysis. Finally, it quantitatively describes the spread process of topics through the number of word frequencies, and analyzes its influencing factors. For Task 2, this paper considers this problem as a typical evaluation problem. Therefore, this paper constructs the AHP analytic hierarchy process model to solve it, and constructs the expert evaluation matrix to score and evaluate it, so as to get the influence of these factors.

Keywords: Text Mining; Jieba; Python; LDA Theme Analysis.

1. Introduction

1.1. Background

Open some information audio-visual APPs, it is not difficult to find that some information with eye-catching titles and no content has been pushed in large quantities. This information is often emotional expression, one-sided interpretation, and some are even presented in the way of spoofing, novelty hunting and vulgar. This phenomenon is closely related to the mechanism of human psychology. There is a "screaming effect" in psychology. For example, in a crowded street, if someone suddenly screams hysterically, it can quickly attract people's attention and attract eyeballs. The "screaming effect" is also prominently reflected in the information dissemination. The vulgar content such as thrillers, spoofs, porn, etc., which is illegally captured, edited, and adapted, often can quickly attract a lot of attention from people. Whether it is to meet people's curiosity, or trigger people's criticism, the communicator can obtain high traffic and click rate from it.

The concept of "information cocoon room" was first proposed by American scholar Thornstein. Through his research on the Internet, he published the book "Information Utopia" in 2006, and stated the concept of "information cocoon room" in the book. He believes that whenever information is spread on the Internet, if people only understand the information and fields they are interested in or choose because of their own information needs, that is, people only listen to what they choose and can please them, Will form "Information cocoon room. In addition, once Internet users form a reading habit, they will be limited to specific fields by their own interests. They are very familiar with the information and topics in this field and their interest is becoming more and more stable. However, they lack attention to the problems and hot spots in other fields, so that people's attention and interest are locked in a smaller range. As time goes by, they will completely seal themselves in a cocoon like

a silkworm Inside the shell, little information is received from the outside. People's understanding of real life began to become blurred, or even distorted.

Based on the above papers, this paper further explores the "screaming effect", "echo effect" and "information cocoon room", to study the transmission mechanism of online information public opinion, and thus intervene in it.

1.2. Our Works

For the above discussion, this paper compiles the program through python software, accesses the microblog port api for crawler operations, and crawls popular topics on microblog. After that, this paper will conduct data processing operations, use the Chinese word segmentation in the jieba database to conduct Chinese word segmentation, word frequency statistics, and lda topic analysis on topics, to conduct public opinion dissemination analysis on topics. The specific flow chart is as follows

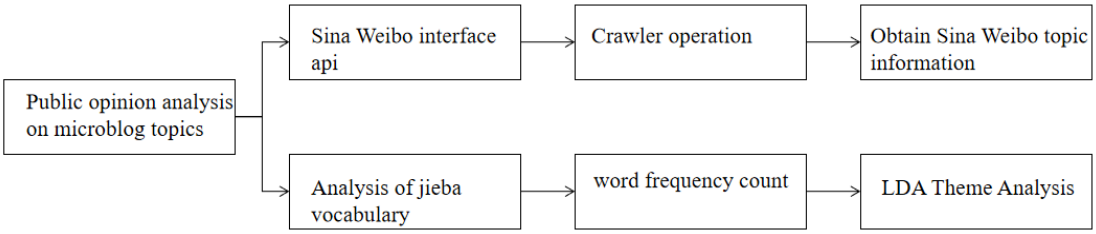

Figure 1. Research Technology Roadmap

2. The Description of the Problem

2.1. Analysis of Specific Issues

2.1.1. Analysis of Task 1

For Task 1, this article uses crawler technology to connect to the Weibo search API interface. The model uses Python's request package and achieves pagination to obtain Weibo data by pulling down the Weibo list. Except for the page parameter that keeps scrolling, all other parameters are fixed and unchanged. From this, the access interface can be determined. Afterwards, this article utilizes the Jieba library for regularization expression and Chinese word segmentation operations, and constructs a word cloud graph for statistical analysis. Finally, the propagation process of the topic is quantitatively described by the number of word frequencies, and its influencing factors are analyzed.

2.1.2. Analysis of Task 2

For Task 2, Regarding question two, the title requires the establishment of a mathematical model to depict the mechanism of neutral consensus and viewpoint polarization, explore the formation mechanisms of "scream effect", "echo room effect", and "information cocoon room", and discuss the influence of factors such as topic attractiveness, user activity, user psychology, mutual influence between different users, and platform recommendation algorithms on the formation of these phenomena. For this issue, this article considers it to be a typical evaluation problem. Therefore, this article solves it by constructing an AHP hierarchical analysis model, and evaluates it by constructing an expert evaluation matrix to obtain the impact of these factors.

3. Assumption

1. Assuming that in this article, all data texts are true and valid.

2. Assuming that in this article, after preprocessing the text data, all the data obtained are true and valid.

3. Assuming that there are no other malicious evaluation data in this article.

4. Symbols

The following table describes the symbols.

Table 1. Symbols and significance

Symbol	Significance
μ	average value
σ	standard deviation
IQR	Interquartile distance
X_{ij}	Data samples
Topic	Number of topics

5. Task 1

5.1. Establishment and Solution of Task 1 Model

This article enters the Google web page review element interface by touching F12, clicking on Net Work, filtering to the XHR tab, and observing the interface returned by the page and the JSON data returned by the response. The results are as follows

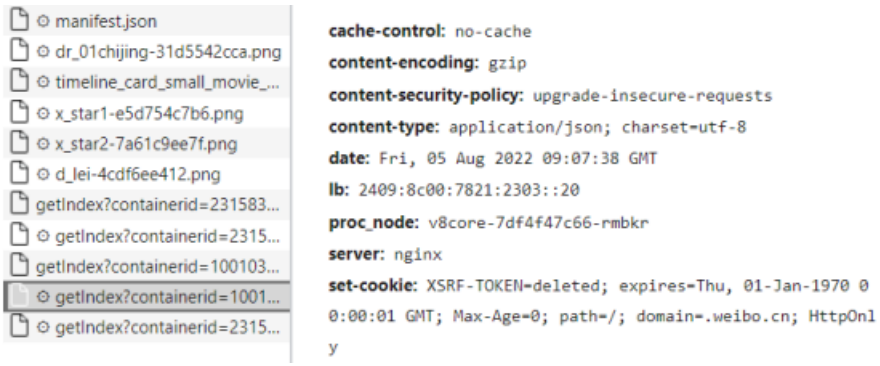

Figure 2. Google Browser Review Element Graph

According to the above figure, the pattern of requesting URLs is as follows: only the page_ The type element is transformed, while the other elements remain unchanged. Therefore, you can use the page_ Modify the type element to perform crawling operations.

Based on the above analysis, modeling and solving were conducted, and the results are as follows

Figure 3. GraphWord Cloud Map Results

By analyzing the word cloud graph in the above figure, it can be seen that words with a large number of words will be magnified. Therefore, based on the above analysis, we can understand the emotional orientation of evaluation users.

Based on this, this article extracts all the topics in LDA and visualizes them using LDAVIS. The two topics with the widest discussion scope in the text content are selected as the evaluation of whether the topics are neutral formulas or extreme expressions.

For the topic of "opposing gender discrimination", the results obtained are as follows

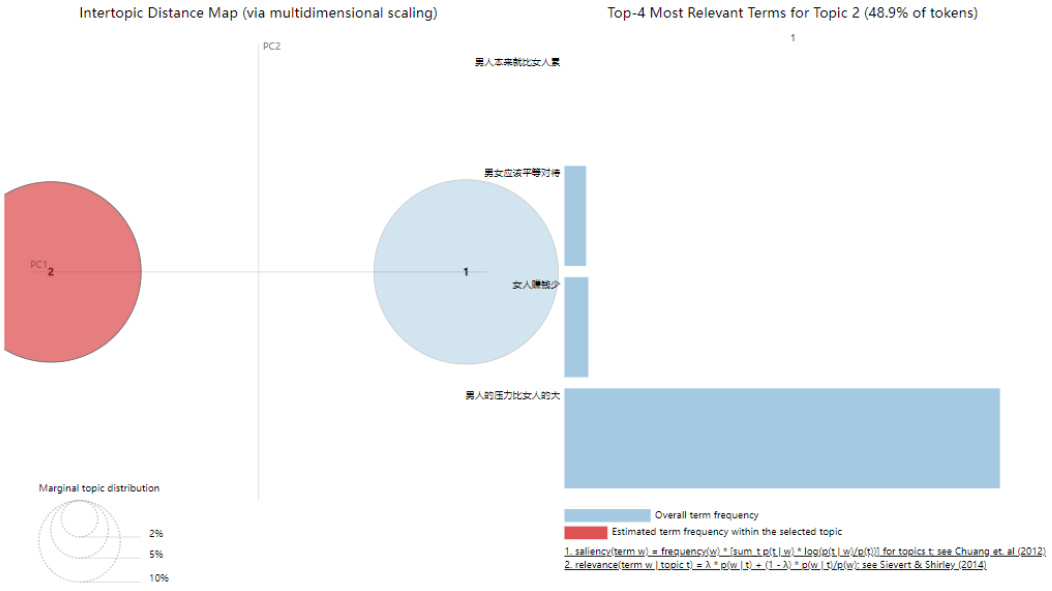

Figure 4. Presentation of the topic of 'opposing gender discrimination'

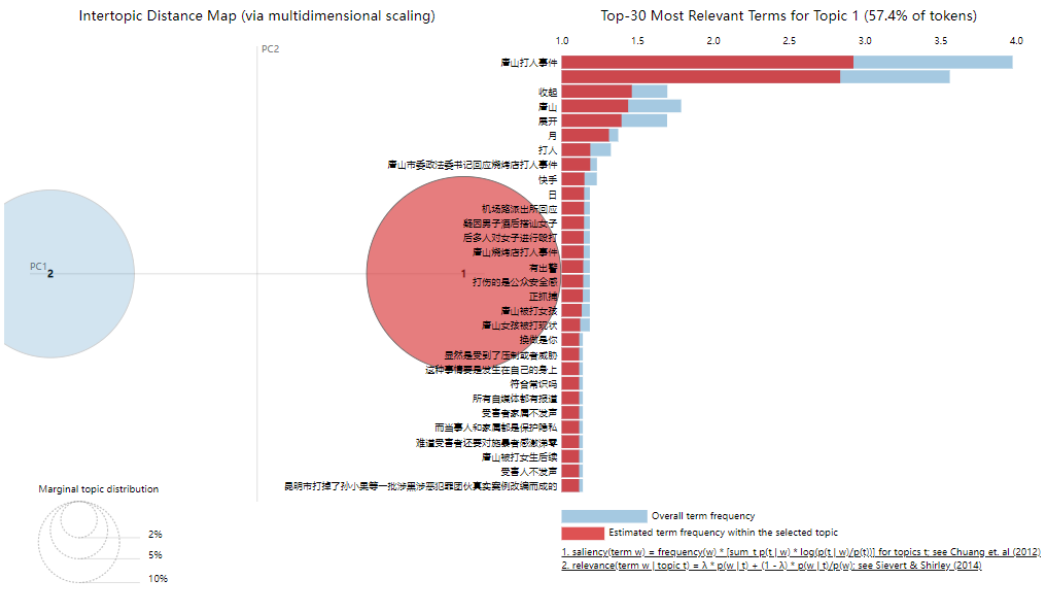

Figure 5. Presentation of the 'Tangshan Hitting Incident' Topic

5.2. Establishment and Solution of Task 2 Model

For Task 2, the title requires establishing a mathematical model to depict the mechanism of neutral consensus and viewpoint polarization, exploring the formation mechanisms of "scream effect", "echo room effect", and "information cocoon room", and discussing the influence of factors such as topic

attractiveness, user activity, user psychology, mutual influence between different users, and platform recommendation algorithms on the formation of these phenomena.

Based on this, the flowchart constructed in this article is as follows

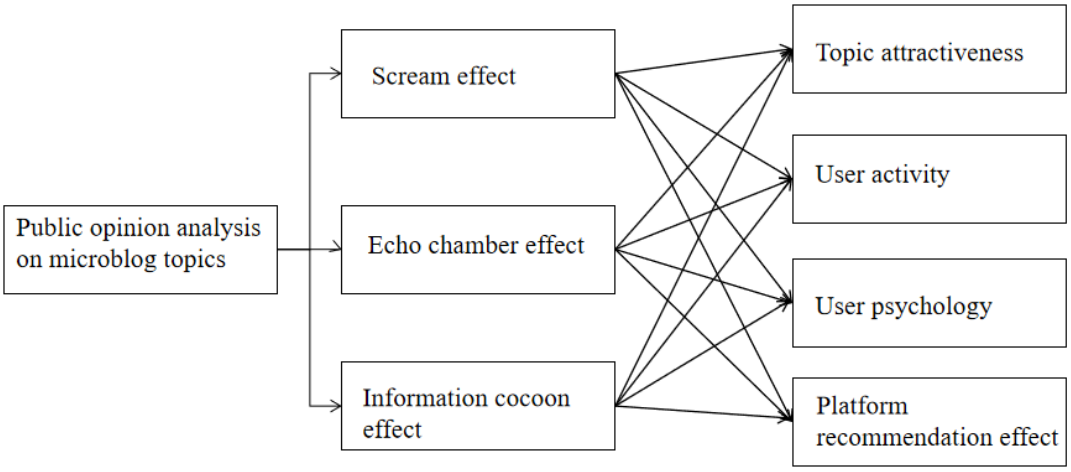

Figure 6. Task 2 Flowchart

For this issue, this article considers it to be a typical evaluation problem. The formation mechanisms of "scream effect", "echo room effect", and "information cocoon room" are highly related to the popularity of topics, the attractiveness of topics, user activity, user psychology, mutual influence between different users, and platform recommendation algorithms. Therefore, this article solves it by constructing an AHP hierarchical analysis model, and evaluates it by constructing an expert evaluation matrix to obtain the impact of these factors.

Firstly, this article introduces the Analytic Hierarchy Process model as follows：The Analytic Hierarchy Process (AHP) model, abbreviated as AHP, was officially proposed by American operations researcher Thomas Saaty in the mid-1970s. The Analytic Hierarchy Process (AHP) model decomposes decision-making problems into different hierarchical structures based on the overall goal, sub goals at each level, and evaluation criteria. Then, by solving the eigenvectors of the judgment matrix, the priority weight values of each element at each level for a certain element at the previous level are obtained. Finally, the weighted sum method is used to gradually merge the final weight values of each indicator at each level for the overall goal. The one with the highest final weight value has the greatest impact on the upper level goal. The so-called "priority weight value" here is a relative measure that indicates the relative superiority of an indicator under a certain evaluation criterion or sub objective, as well as the relative importance of each sub objective to the upper level objective.

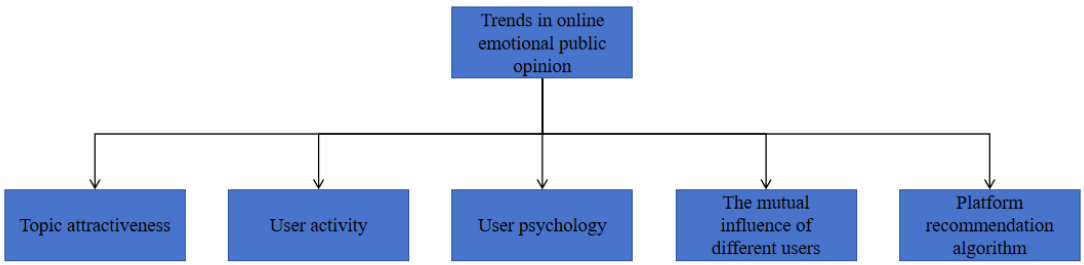

Figure 7. Analytic Hierarchy Process Network Structure Diagram

Based on this, this article constructs a 5-dimensional comparative matrix, which is constructed as follows

According to the analysis of the data in the above figure, it can be seen that among the various indicators of "scream effect", "echo effect", and "information cocoon room", the platform

recommendation algorithm has the largest influencing factor, followed by topic attractiveness and the influence between different users, and the lowest is user activity and user psychology.

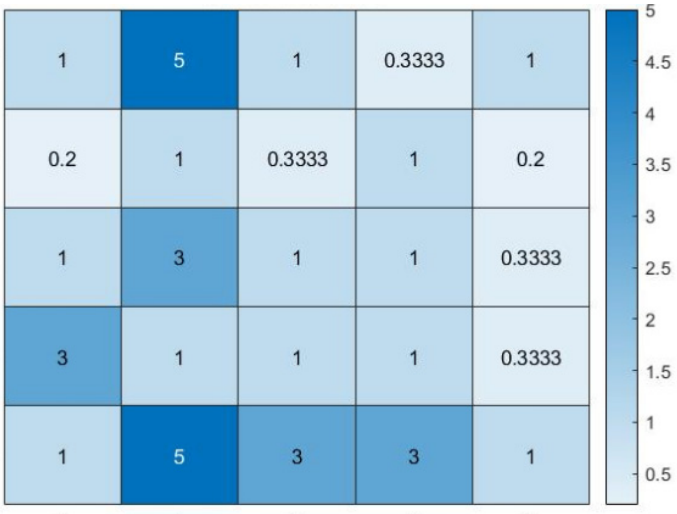

Figure 8. Correlation thermogram

By analyzing these impact indicators, it can be seen that the platform recommendation algorithm determines the way and quantity of user observation information, forming a "cocoon of information" to a certain extent, resulting in the "echo effect", so the influence of this factor is the greatest.

Secondly, due to the fact that the attractiveness of the topic and the influence between different users to a certain extent determine the popularity of the topic and the degree of offline discussion among associated users, the degree of impact is relatively high.

For user activity and user psychology, the correlation between user activity and whether there is an "information cocoon" is relatively low. This indicator is only a feature that indicates that users are more likely to see certain topics, so the impact is not significant. Therefore, by analyzing the above results, it can be seen that the results obtained through the AHP hierarchical analysis model are consistent with the actual results, so this model can effectively detect the degree of influence between various factors.

6. Conclusion

For the top-level design of the government, as a public institution, it plays a management role in the direction of public opinion. Therefore, for the top-level design of the government, it is recommended that the government should control the direction of public opinion, play a role in macro control and micro regulation, and timely reverse some misguided public opinion, playing a regulatory role.

As a mainstream media, we should do a good job in public opinion promotion. Because within a nation, there is a mainstream consciousness of social values, which is called "mainstream values". In today's highly developed media and information overload, many users fail to pay appropriate attention and understand information in the process of receiving it. So, spreading mainstream values through the internet can not only maintain social order, but also play a decisive role in various types of information. There are many public information platforms in China, such as Weibo, with a large fan base. Therefore, mainstream media should publish authoritative and credible information. Mainstream media such as "People's Daily" and "CCTV News" have a large number of fans, and should adhere to professional and social responsibilities in promotion, and publish with caution; Enable netizens to gain trust, demonstrate their authority, and lead them to break through the "information cocoon".

As a general user, users have a dual identity as information dissemination recipients and participants in the internet, and are also the main body of group polarization behavior. For individual users, they should actively shape their cognitive breadth and avoid falling into the "information cocoon", let

alone participating in group polarization behavior. As recipients of information dissemination, users in the early stages of using Weibo will follow their accounts based on their personal interests and preferences. Over time, their followers gradually become stable, and the required amount of information gradually becomes saturated. This will lead to a fixed channel for information acquisition, overlapping similar information, and limited cognitive range. Therefore, in order to avoid being trapped in a cocoon, we need to adjust our following accounts in a timely manner, constantly refresh our fixed followers, and "disconnect" some information; As participants, one should enhance their personal media literacy. When engaging in online communication, one should correctly recognize the virtual and fragmented nature of online information. When self-expressing online, one should pay attention to wording to avoid ambiguity. When encountering intense behavior or extreme speech attacks, one should speak rationally and retreat appropriately if necessary. Online social media platforms are not places to judge personal value, only shallow and deep exploration can provide a better information experience. When communicating online information, constantly shaping one's own cognitive breadth of information and improving one's personal media quality can truly, comprehensively, and objectively disseminate news information, and create a healthy and good online space.

7. Evaluation and Promotion of Model

7.1. Advantages of Models

1) This article combines MATLAB software and Python software for programming, and the obtained data is preprocessed in advance, resulting in authentic and reliable results.

2) The model constructed in this article is based on text information mining, and extensive research has been conducted on comments on Weibo. The results obtained are in line with the actual situation.

7.2. Disadvantages of the Model

The assumption of the model in this article is too idealized, and some practical situations should be considered to have an impact on the model.

References

[1] Liang Feng. Information Cocoon Room [J]. News Sentinel, 2013 (01): 87.

[2] Research on Short Video Communication Mechanism Based on Immersion Theory [J]. Li Peihan. News Culture Construction. 2022 (02).

[3] The "Time Paradox" of Information Wealth Differentiation: An Empirical Test Based on the Boundary Elements of Personal Information World [J]. Zhou Wenjie, Bao Yunqi. Library and Information Knowledge. 2021 (06).

[4] Knowledge Acquisition, Artificial Intelligence, and Library Spirit [J]. Jiao Yanpeng, Liu Wei. Journal of China Library. 2021 (05).

[5] Analysis of "Traffic Landscape" - Taking the "Xiao Zhan Incident" as an Example [J]. Liu Yixuan. Journal of Gansu Radio and Television University 2021 (04).

[6] The Causes and Countermeasures of "Chaos" in the Pink Circle from the Perspective of Communication Science [J]. Zhang Yixin. New Chronicles. 2021 (13).

[7] The negative impact and breakthrough of the "information cocoon" effect on ideological and political work in universities [J]. Liu Zefeng, Liu Ming. Journal of Shandong Youth Political College 2021 (02).

[8] Research on the Formation Mechanism and Guiding Strategies of Group Polarization on Social Media Platforms [J]. Lu Zhizeng, Gao Xiang. Intelligence Theory and Practice. 2021 (08).

[9] Reflection on the Study of "Information Cocoon Room" in Fragmented Communication: Based on Context Transformation, Cocoon Room Value, and Subjective Initiative [J]. Chen Chunyan. News Research Guide. 2021 (03).

[10] How can college students break through the "information cocoon" in the era of intelligence--From the Perspective of Information Literacy Cultivation [J]. Zhang Min, Wang Pengjiao, Meng Xiangyu. Modern Education Technology. 2021 (01).