

"Internet Witch Hunt" is the Sharp Edge of the Internet Era

-- Based on Python Analysis of the Public Opinion and Emotional Analysis of "Online Witch Hunting" by Mainstream Media and Online Social Media

Yi Xue

Shanghai University, Shanghai, China

Abstract. In the Internet era, personal freedom of speech has been fully reflected, but it has also brought a lot of negative effects. "Internet witch hunting" refers to human flesh search, information collection, personal attacks on specific people on the Internet, and stirring people's emotions through the Internet. In recent years, with the development of the Internet, Internet rumors and online violence and other phenomena began to spread on the Internet. On online social media platforms, personal freedom of speech has been fully reflected, but it has also brought great hidden dangers to public security. With the development of technology, personal privacy has been paid more and more attention. How to protect personal privacy and prevent network violence has become a common problem in today's society. Starting from Python language, this paper will analyze the content of word frequency, word cloud map, theme modeling, emotion analysis results of mainstream media and social platforms for "online witch hunt" events, so as to provide a certain reference for us to better protect the personal privacy of ourselves and others.

Keywords: Data Mining; Data Analysis; Internet Witch Hunt; Sogou News; Weibo; Bilibili.

1. Foreword

In Europe from the 15th and 17th centuries, the prey movement was in full swing, with about 200,000 people being investigated, abused and even killed because of their identified as witches. As a pragmatic manual of medieval Christian witch hunt, Witch Mallet is a true record of how to demonstrate the existence of evil, and how to detect, try and execute witches. However, "witch hunting" is ignorant and absurd, a kind of ritualized social violence. It was not until the "Enlightenment movement" brought the ideological liberation of rational thinking, democracy and freedom that people fled from the dark world of "witch hunting".

Now, in the 21st century, the hunt for witches has disappeared, but another form of "witch hunting" has grown on the Internet. Both the "200 yuan takeaway red envelope incident" during the epidemic period and the dirty language in the comments section of various online platforms all reflect the reappearance of the "witch hunt movement" on the Internet. People often focus on attacking the selected targets, and forcibly interpret their behavior and discourse, so that they are violent from all aspects of psychology and life.

What causes the "Internet witch hunt" in the Internet age? Why did the lines of the beating keyboard become "death notes"? What is the psychology behind these "witch hunters"? This paper will analyze the word frequency, word cloud map, theme modeling, emotion analysis results of "network witch hunting", as well as feature analysis, cause analysis and solution exploration, so as to show a more detailed and in-depth understanding and analysis of "network witch hunting".

2. Problem Definition

First of all, the scope of "Internet witch hunting" studied in this paper is mainly defined in articles on the "mainstream media platform", as well as bilibili and Weibo, two popular online social media. Choose the three platform is that the three platforms (1) the traffic, able to climb more sufficient data (2) high credibility, gave the more credible data and articles (3) representative, can represent the current surfing the young people for popular events (4) the difference is big, the mainstream media is

serious, bilibili the video site focuses on information transmission, weibo can widely express inner feelings, the organic combination of the three can help us more comprehensive view of "network witch hunt" events.

So this article will revolve around: when is the hot time for the "Internet witch hunt"? Why is it so popular in these times? What are the key words and content of "Internet witch hunt"? What are the themes of "Internet witch hunt"? What kind of emotional attitude do users on mainstream media and mainstream social media treat "Internet witch hunting"?

Secondly, the main definitions of the problems in the emotion analysis through Python language are as follows:

1. The emotional classification of the "online witch hunt" event in this paper mainly studies the emotional analysis results of the event in the mainstream media and online social media;
2. The data set used in this article mainly includes news reports on "online witch hunting" events published by mainstream media websites such as CCTV News, Sina Weibo and People's Daily Online and related articles published by online social media accounts, with a total of 376 articles.
3. This paper uses Python language to analyze the data of media and social platforms. The main technologies include: text pre-processing, feature extraction, bag of words model, TF-IDF, etc.
4. This paper mainly analyzes the reports and emotion classification results of "witch hunting" online in mainstream media and social platforms through crawler technology and text preprocessing technology.

3. Analysis Objects and Data Sources

The data of this article comes from sogou News, a mainstream news and information platform, and mainstream social media platforms such as Weibo (more than 100 information), bilibili (more than 40 information) and Sogou News (more than 40 information).

The data analysis methods used in this paper include: pandas (data processing), jieba (word segmentation), wordcloud (wordcloud), and sklearn.feature_selection (feature selection), sklearn.metrics (Metric index), sklearn.cluster (Cluster), gensim (Extract semantic themes), snownlp (emotion analysis), pyLDAvis (theme model visualization) and visualized keywords.

When sorting out the data, the author can preliminarily see from the data that most of the media have paid attention to this event, and have paid some attention to the personal privacy protection issues involved in this incident. In addition, mainstream media including Sina, Sohu, netease and Tencent also released relevant reports in the first time to provide public opinion support for the incident.

However, due to the particularity of this event, such as privacy leakage and other issues involved, the data analysis results of media reports and social platforms need to be further verified.

In addition to media coverage, social media platforms will also comment on such events. The author also included the comment data of social platforms in this analysis to understand the social platforms 'views on such events and their attitudes towards users' comments. These comments can also be used as a supplement to the media coverage.

In addition, through the analysis of the reports of major news portals in the past year, it is found that most news portals use keywords such as "human flesh search" and "Internet witch hunt" when reporting this event, so the attitude of the media towards this event should also be taken into account in the emotional analysis.

4. Data Preprocessing

First, the raw data of the "Internet witch hunt" event is obtained in Python, then converted into Chinese text data, and finally stored in a string form. Use the `set_text` function in Python to get the time stamp of the text, and then do a date format conversion to convert it to a one-dimensional string.

Since the data set contains relevant information about the network social platform, the `requests` library is used in Python to obtain the name and address information of the social platform. With the `requests` library, we can obtain the social platform name and save it to the variable `requests_text`.

Use the `pandas` library to get the date format string in Python and save it as a string.

After making the word cloud map and analyzing the word frequency, the author sorted out the popular time line of "Internet witch hunt", and finally used TF-IDF statistical method to extract the keywords of "Internet witch hunt" events, conducted text lineage clustering with `scipy`, and built the LDA theme model and visualized and analyzed it.

5. Data Visualization and Analysis

1. Frequency of word analysis of "Internet witch hunting" in main stream media and social media

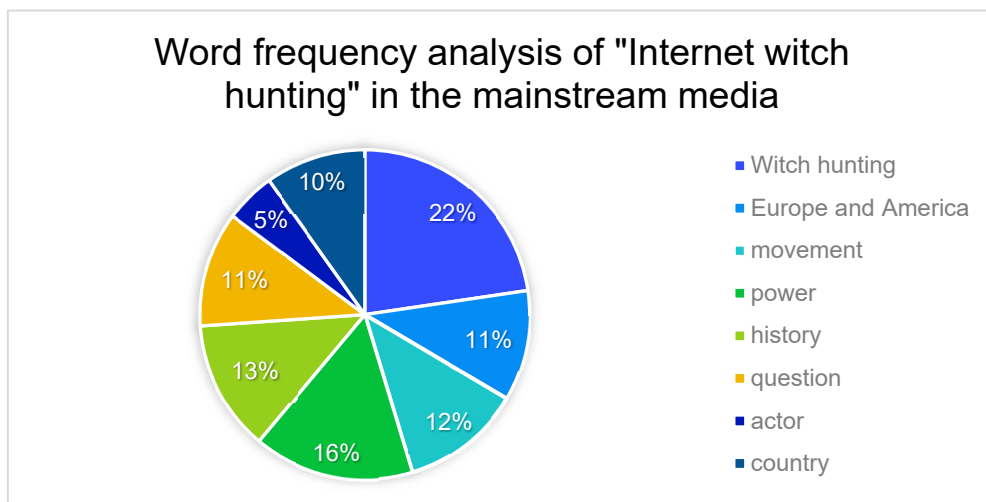


Figure 1. Word frequency analysis of "Internet witch hunting" in mainstream media

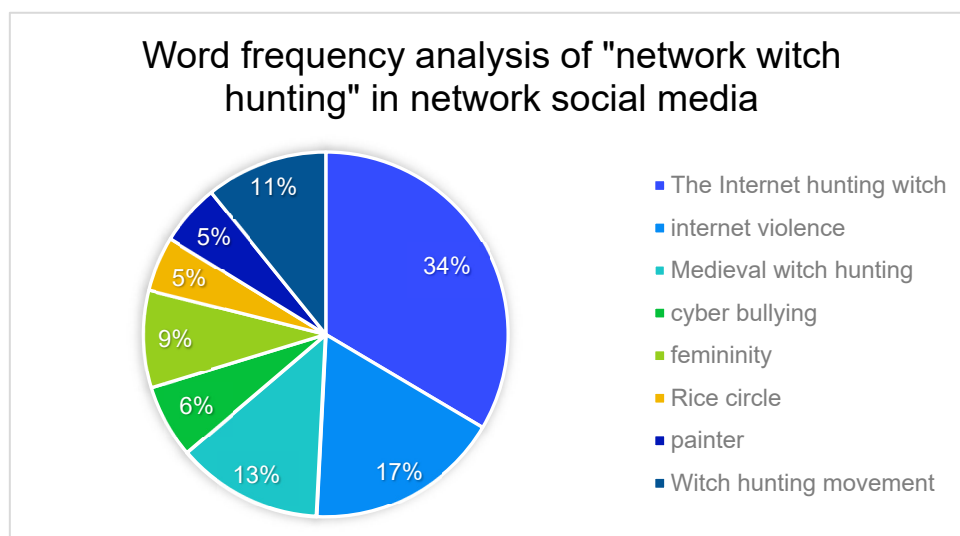


Figure 2. Word frequency analysis of "Internet witch hunting" in network social media

Through word frequency analysis, you can know which words in the text the frequency is higher, and then help better understand the text content and theme, after the data collection and cleaning, the text data in accordance with the certain rules, such as space or punctuation as a separator, and then statistical word frequency, according to the word results, the number of each word in the text, and in accordance with the order from high to low, finally the author choose through the pie chart presented visual results.

As shown in the mainstream media in the emergence of the word "Internet witch" ratio than the largest, close to a quarter, and the difference between the keywords in the mainstream media more tend to the macro sense of political culture problems, such as "power", "history", and keywords in network social media to social popular events, such as "painter", "rice".

It can be seen that both mainstream media and online social media focus on the "online witch hunting" itself and related events, without ignoring their historical reasons and social events, but both have different emphasis and seriousness.

2. Word cloud map analysis of "Internet witch hunting" in mainstream media and social media



Figure 3. Cloud map of "online witch hunt" words related to mainstream media and online social media content

Since this section focuses on the analysis of the content of "Internet witch hunt", and "Internet witch hunt" itself has a certain topic heat, but its spread scope is still not wide enough, the author chooses to combine the content of Weibo, Sogou News, bilibili and Weibo with an Exel form for word cloud map analysis.

Through the word cloud map, we can see that "network" and "witch hunt" are the most obvious, which shows that the "witch hunt movement" is now constantly "network", and it is also given more meaning. For example, "artist", "985 student", "artist" and other ID cards are being added, while "woman", "woman" and "girl" are even more hurt by "Internet witch hunting". There are also some words "rumors", "violence" and "cruel", reminding us of the violence of "Internet witch hunting".

3. The "Internet witch hunt" hot time period

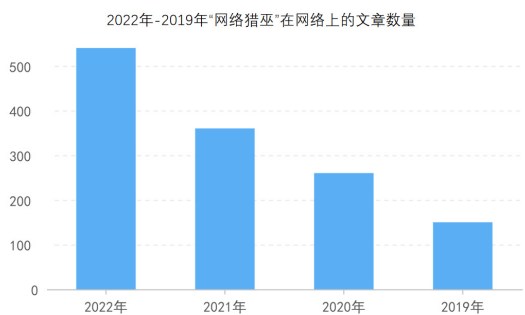


Figure 4. Bar chart of the number of "Internet witch hunt" on the Internet in 2019-2022



Figure 5. Take 2021 as an example of "Internet witch hunt" year-heat line chart

Table 1. Inventory of the key events of "Internet witch hunt"

event	Cause mental harm to the parties concerned	Cause material injury to the parties concerned	The negative impact on the network public opinion environment	No formal result or apology statement
2020.7 The woman was rumored cheating on the Courier	√	Unemployed, depressed	√	The rumour-monger was sentenced and apologized in court for 10 months
2021.3 Chengdu university iPad witch hunt incident	√	×	√	×
2021.5 Female singer Naiwan suffered a net storm	√	×	√	×

As we can see from the figure above, "Internet witch hunting" has attracted more and more people's attention over time. From 2019 to 2022, articles on Internet witch hunting on the Internet have significantly increased, and the popularity of "Internet witch hunting" events will also be strengthened and enhanced with the explosion of some social news events.

This also reflects from the side that the today's network world exposes the lag of China's current supervision system and the imperfect moral norms. There is no clear legal provision on the scope of Internet privacy and freedom of speech, and the right definition of netizens is still pending. The national supervision and guidance also need to be improved and strengthened, and cyber violence will cross the bottom line of morality and break social norms at any time.

"Witch hunter" often after a long wait and struggle to get due compensation, and events intensified process, many ignorant people also unconsciously involved in the "witch hunter" process, they in the absence of knowing the truth because of "information" and "collective unconscious" thought "justice" to participate in the network violence. Therefore, in order to get out of the dilemma of network witch hunting, we should emphasize the improvement of the national legal system and the platform audit.

6. Theme Modeling Analysis of "Internet Witch Hunt"

After constructing the dictionary, eliminating the high frequency words and constructing the corpus, 10 themes were generated through the LDA model study, and each theme contained 30 words. Then the author visualised the LDA theme model, and the results are as follows:

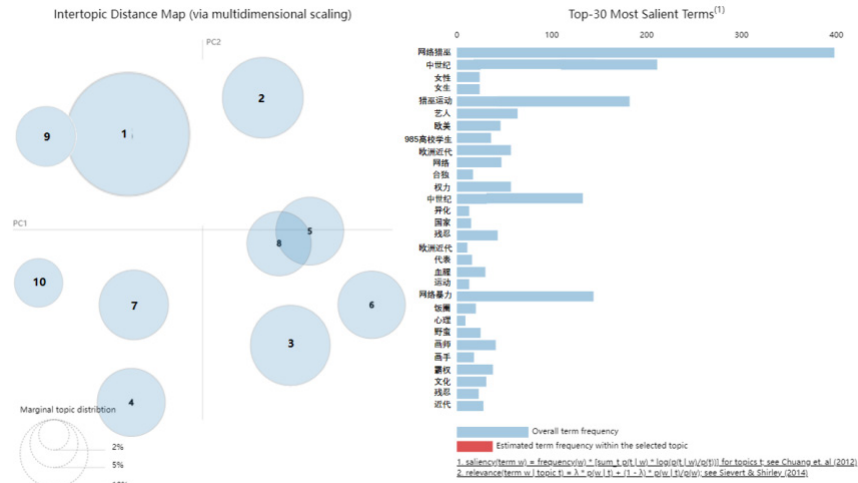


Figure 6. Theme modeling of "Internet witch hunt"

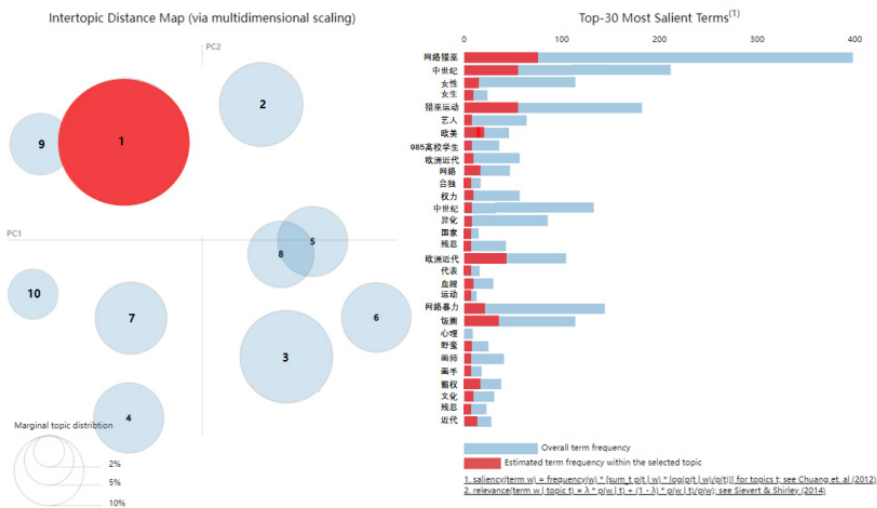


Figure 7. Theme modeling of "Internet witch hunt" 1

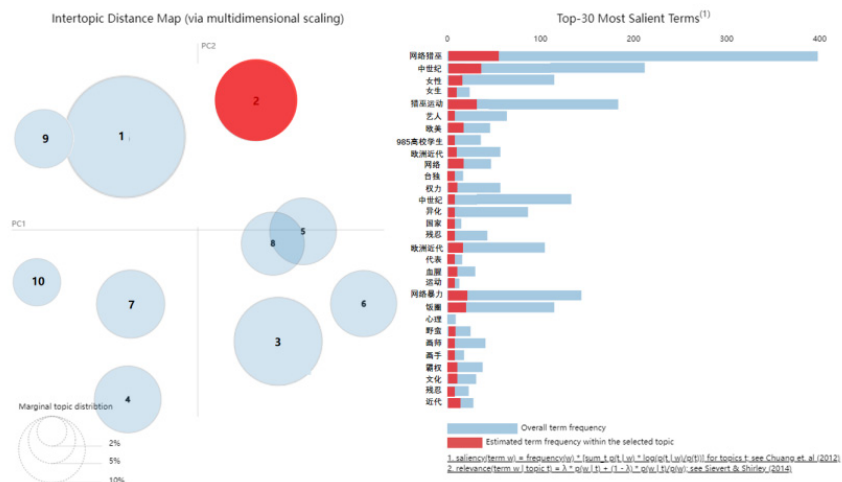


Figure 8. Theme modeling of "Internet witch hunt" 2

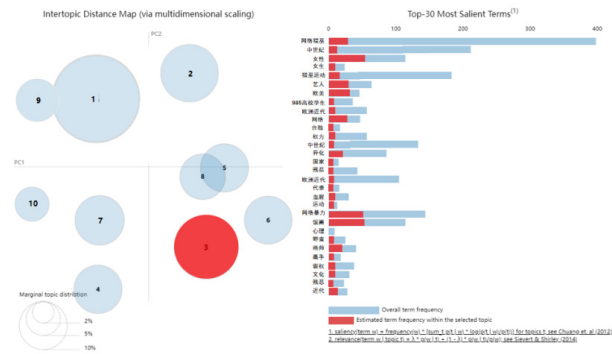


Figure 9. Theme modeling of "Internet witch hunt" 3

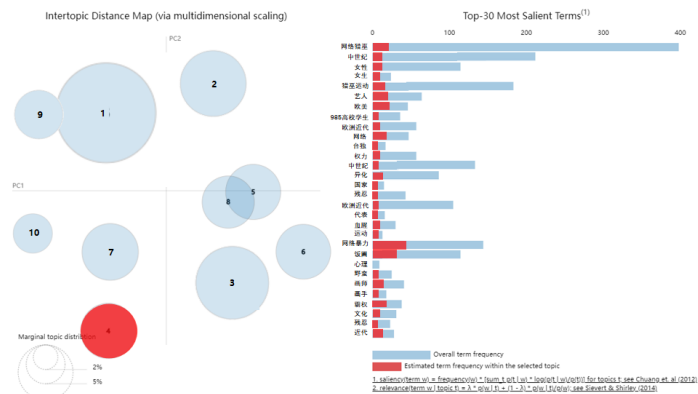


Figure 10. Theme modeling of "Internet witch hunt" 4

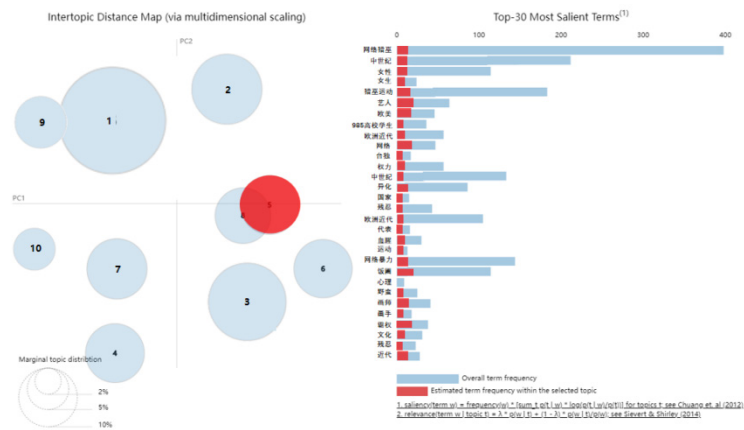


Figure 11. Theme modeling of "Internet witch hunt" 5

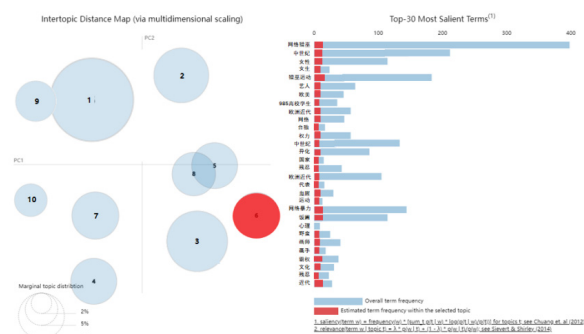


Figure 12. Theme modeling of "Internet witch hunt" 6

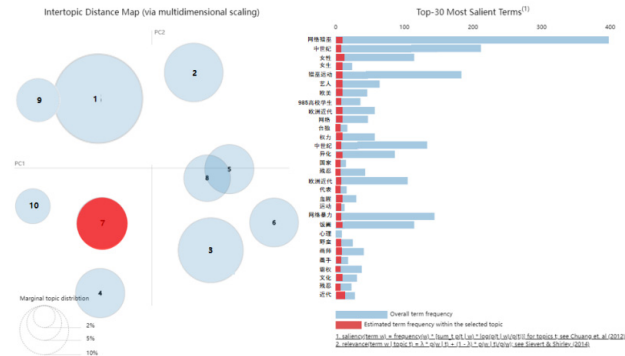


Figure 13. Theme modeling of "Internet witch hunt" 7

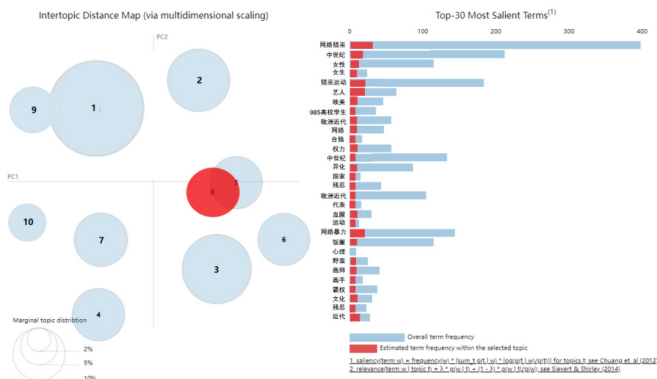


Figure 14. Theme modeling of "Internet witch hunt" 8

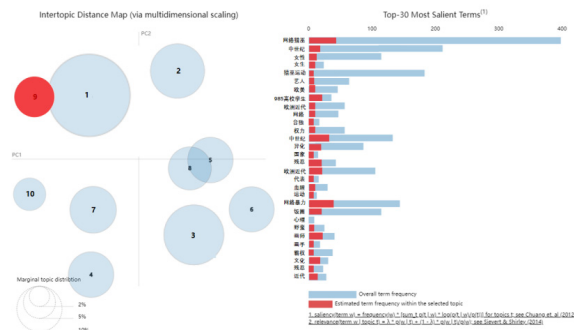


Figure 15. Theme modeling of "Internet witch hunt" 9

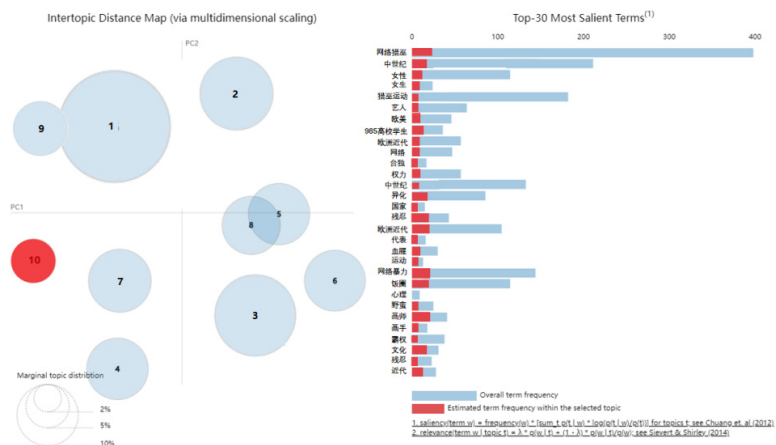


Figure 16. Theme modeling of "Internet witch hunt" 10

In the pyLDAvis output, each bubble on the left represents a theme, and the corresponding words and bars on the right side of the bubble are the significant keywords constituting the selected theme, and the larger the bubble, the more common the theme becomes. By the figure, the theme of "network witch hunt" can be roughly divided into three parts: (1) the Internet witch hunt itself, on behalf of the keywords for "Internet witch hunting", "witch hunt", "network" (2) the embodiment of related hot events, representing keywords for "painter", "artist", "985 students" (3) history related prey events, on behalf of the keywords for "history", "Europe", "medieval".

And (3) in the keywords often accompanied by (1), is to explain the reason of "network witch hunt" and historical origin, visible "network witch hunting" visibility and interpretation in constantly rising, in (2), the number of the most extensive and complex, some unrelated topics may be involved because "network witch hunt", such as "ipad witch hunt events" will "ipad" the unrelated concept and "network witch hunt" tied together.

(3) Emotional analysis of "online witch hunting" between mainstream media and online social media "Sentiment" represents emotional polarity, the value range is $[-1, 1]$, -1 represents completely negative, 1 represents completely positive; Subjective represents the degree of subjectivity.

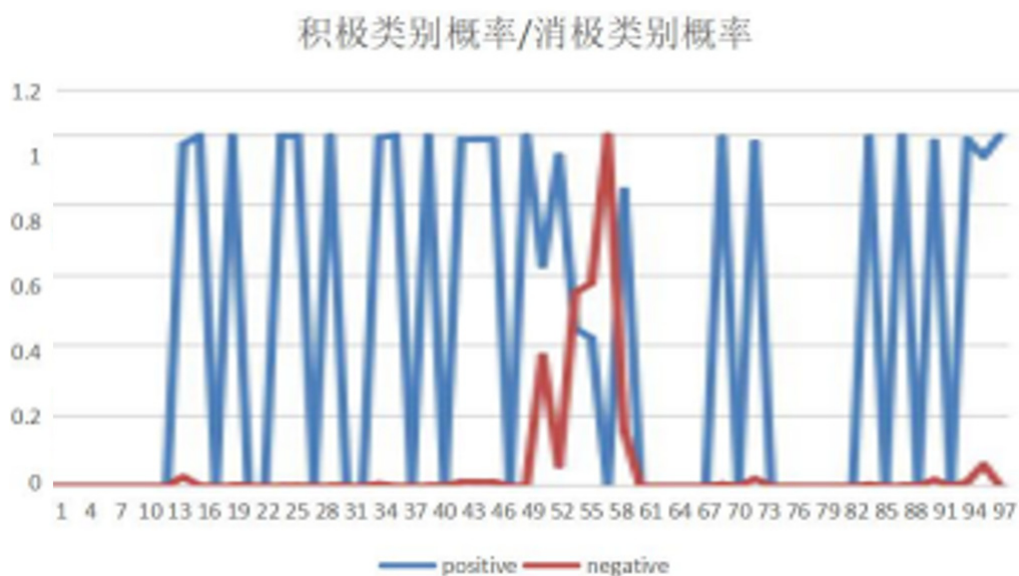


Figure 17. Analysis of the positive and negative emotions of "online witch hunting" in mainstream media

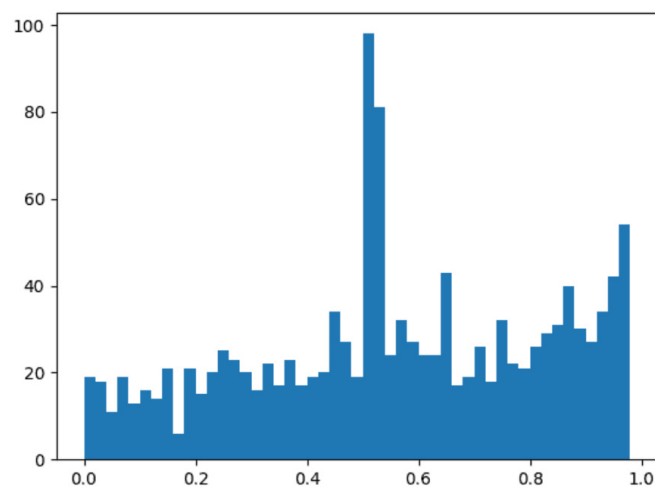


Figure 18. The distribution of emotional value of Bilibili platform users

Through Python's natural language processing technology, we analyze the emotional attitude of mainstream media and social platforms in the "Internet witch hunt" event, and draw the following conclusions:

- (1) The mainstream media has a neutral and positive attitude towards the incident
- (2) The emotional expression of social platforms generally shows positive emotions, among which Weibo users express a neutral attitude, while bilibili users show a positive emotion.
- (3) For victims, online witch hunting may bring huge negative emotions, such as fear, anger, helplessness, etc., which are often amplified and quickly spread to a wider audience.
- (4) For the participants of online witch hunting, some people may have the needs such as comparing and grabbing attention, while others may participate out of the identification of social justice and other values, and their emotional tendencies may be different.
- (5) For the media reporting on the online witch hunt events, they will also be affected by the pressure of public opinion and the emotional comments, and the emotional colors presented by the reports may be different.

Therefore, when doing the emotional analysis of the network witch hunting, it is necessary to take into account various factors and perspectives, and with the help of natural language processing and other technical means, to excavate the emotional image, emotional polarity and other information behind the text, so as to better understand the nature and influence of the network witch hunting phenomenon. However, it should be noted that in the analysis of online witch hunting, we should abide by the network ethics and laws and regulations, and should not conduct personal attacks, rumors and fabrication and other illegal acts.

7. Analysis of the Results and Conclusion

We summarize the following points from the comments on the "online witch hunt" incident on major mainstream media and social platforms:

The influence of "Internet witch hunt" event is relatively wide and involves a large number of people, but it still needs to improve the attention of some "passers-by", which can increase the cognition of ordinary people about "online witch hunt" event to some extent. Some of the comments on the "online witch hunt" incident are highly aggressive and personally emotional. After being exposed, there were a large number of netizens denouncing and abuse under the related topic, which also shows from the side that the "network witch hunt" event is of great harm.

In the "Internet witch hunt" event, the emotional reaction among netizens is relatively intense. It is mainly manifested in the personal attacks on the relevant parties, human flesh search, and malicious rumors. Most netizens expressed their sympathy for the incident, and also hoped that the relevant departments could investigate the truth as soon as possible and give the public an explanation. Some netizens also said that they disagree with these remarks, thinking that the "Internet witch hunt" incident is just a farce. I hope everyone can treat this matter rationally. In a word, for the "Internet witch hunt" incident, we should strengthen the norms from the laws and regulations.

8. Analysis and Future Outlook

In the era of the Internet, the beating keyboard may hit the next line of "death list", behind the seemingly small words, it is endless network violence attacks, but also a net of ignorance, one-sided prejudice, passion and backwardness. "Network witch hunt" is a blade of the Internet age, it will white for black, it will justice stigmatized, it will gather a terrible group force "hunting", however, I believe that in the public to realize its lethality, realize its irrationality, understand its principle, through personal, social and media three level change also the Internet a moral clarity.

With the continuous development of technology, "Internet witch hunting" and other problems are getting more and more attention. How to better protect our personal privacy and prevent the problem of "Internet witch hunting" is a common problem faced by today's society.

By analyzing the text using natural language processing technology, the information behind the events can be mined from huge amounts of data. Machine learning technology can be used to analyze the hidden emotional tendencies behind events, so as to identify and combat network violence and other behaviors. But these technologies are still in the development stage, and there are still some problems. For example, at present, there are some tools for the use of machine learning models for emotion analysis on the Internet, but these tools can not solve the problems faced by artificial emotion analysis well, and there are problems such as insufficient protection of personal privacy.

Secondly, in the process of continuous technology development, personal privacy protection will also face great challenges. Due to the growing number of Internet users and the advent of the era of big data, the risk of personal information leakage will also increase increasingly. How to ensure the security of user information has become one of the important problems facing the current Internet industry. With the development of blockchain technology and artificial intelligence technology, the encryption and processing of personal information has become an inevitable trend. If personal information can be encrypted in data collection, storage, use, and exchange links, personal privacy protection will be further improved.

References

- [1] Li Yingfeng. Chinese and English bilingual lost and found being caught by Internet violence is a reverse education class [J]. Shi Dao, 2021 (04): 18.
- [2] Sun Jian, Xu Zuying. Network public opinion supervision and its norms [J]. China Administration, 2011 (12).
- [3] Dai Jintao. The anced protection of privacy and freedom of speech from "human flesh search" [J]. Law, 2008 (11).
- [4] Yu Xiucui. The behavior and motivation of the network public opinion field communication [D]. Huazhong University of Science and Technology, 2011:24.
- [5] Wang Mingke. Witch Crisis: A Thoughts on COVID-19 [J]. Journal of Central South University for Nationalities (Humanities and Social Sciences Edition), 2020,40(03):1-7. DOI: 10. 19898/j. cnki. 42- 1704/c.2020.03.002.
- [6] The blood reflected behind the death of a Taiwanese female writer [J]. Sound, 2017 (10): 44.
- [7] Xu Shanwei. Why women became the main victim of the witch hunt movement in early modern Europe [J]. Historical research, 2015 (05): 157-172 + 193.
- [8] Research on the psychology of the masses (Gustav Le Pen).
- [9] Li Baihan, Zhou Morfan, Cui Chunyu, Gao Kai. Simple text sentiment analysis based on Python [J]. Science and Technology Education in China, 2022 (03): 62-63.
- [10] Xia Yuqin, single Xue. Simple text sentiment analysis based on Python [J]. Yinshan Academic Journal (Natural Science Edition), 2018,32(04):58-62.DOI:10.13388/j.cnki.ysajs.20180917.001.