Research on Portrait of Online Public Opinion Subject Based on Big Data of Public Opinion——A Case Study of Notre Dame

Nan Jia, Yixue Xia*, Ling Jia
Academy of Intelligent Police, China People's Police University, Langfang 065000, China
shujukexue2020@126.com

Abstract: [Purpose/Significance] As online public opinion gradually becomes an important carrier of public opinion expression, multi-dimensional portrayal of each subject of public opinion based on public opinion big data helps to understand the characteristics of each dimension in an all-round way, thereby enhancing the pertinence of network social governance. This article aims to actively explore the subject of online public opinion subject based on public opinion big data and machine learning algorithms. [Method/Process] Use web crawler technology to crawl the microblog and comment data of the "Notre Dame" event, and describe it from the three dimensions of the basic attributes of microblog users, the microblog comment stand and the topic of microblog comments. In the basic attribute dimension, the reputation index of each user is calculated using information such as the number of fans and followers; in the Weibo comment stand dimension, machine learning algorithms are used to classify the comment stand, and the oversampling technique is used to oversample the data set; in the microblog comment subject dimension, the LDA model is used to divide the subject of the review data; finally, a fusion portrait is developed based on the above portrait dimension. [Results/Conclusion] Through the integration of portraits, comprehensive use of image information such as microblog user influence, comment stances, comment topics, comment time, etc., to propose guidance strategies for different scenarios, helps to improve the targeted and effectiveness of network social governance.

1. Introduction
Currently, online public opinion has become an important vehicle for public opinion expression. Massive, multi-source, fast-updated, and highly relevant public opinion big data provides data resources and technical methods for comprehensively grasping the trend of online public opinion and improving the pertinence and effectiveness of online social governance. Multi-dimensional portraits of online public opinion subjects based on public opinion big data will help provide rapid, comprehensive and accurate decision support for online social governance.

With the rapid development of big data and artificial intelligence technology, user portraits have become an important technical means to analyze the characteristics of users[1]. It can perfectly abstract a full picture of a user’s information for further rapid and accurate analysis of user behavior and habits. Important information such as, opinions, attitudes, etc., provide a sufficient data basis[2]. The landing field of user portraits mainly focuses on accurate recommendation in the fields of business and public services. The research on user portraits in the field of social networks mainly focuses on three themes. The first is to focus on the characteristics and classification of user groups to promote the accurate
push of network information services. For example, Lin Yanxia et al. [3] established user portrait models to divide Weibo users into groups, and carried out precision marketing based on the characteristics of each group. And services provided a basis; Zhao Shuguang [4] used user portrait technology to classify users with high conversion rates and analyzed group characteristics from multiple angles in order to improve the effect of social media operations; Wei Mingzhu [5] and other users Portrait technology studies the characteristics of social media high-impact users from multiple dimensions. The second is to study the emotional characteristics of netizens or understand their wishes. Scholars such as Ye Guanghui, Ren Zhongjie, Deng Xiaoyi [6-8] used user portrait technology to study the emotional characteristics and evolutionary laws of netizens, and provided a reference for grasping the emotional state of netizens in certain situations; Fan Zhe [9] for understanding numbers Indigenous people’s willingness to adopt social media, using user portrait methods to conduct in-depth research on their phase characteristics in the process of adopting social media. The third is to study information dissemination behavior and public opinion guidance measures by relying on social media user portraits. For example, An Lu et al. [10] conducted user portraits on Weibo information and comment users to improve the guiding effect of anti-terrorism public opinion, and compared the portrait results; Liu Haiou et al. [11] constructed users based on online social activity information The portrait model uses the results of the portrait to analyze the user's information dissemination characteristics in depth, which provides an important reference for exploring information dissemination behavior.

Although some scholars have conducted social media user portrait exploration around public opinion guidance in the above research, each portrait dimension is relatively independent and lacks integrated analysis. Therefore, the pertinence and effectiveness of the public opinion guidance strategy developed need to be enhanced. On the basis of existing research, this article is mainly oriented to network social governance, based on the big data of network public opinion, explores a comprehensive and overall portrait of online public opinion subjects, focusing on the opinions and attitudes of public opinion subjects. Therefore, this article proposes a framework for the integration of public opinion subjects, which integrates the characteristics of public opinion from the three dimensions of basic attributes, Weibo comment stance and Weibo comment theme, and achieves a panoramic depth portrait. In the dimension of Weibo comment position, oversampling Processing solves the problem of unbalanced user position label distribution, improves the overall portrait effect and portrait accuracy, and the portrait result helps to improve the pertinence and effectiveness of network social governance.

2.Construction of the Model of Internet Public Opinion Subject Portrait

User portrait generally includes four links: data information collection, user dimension screening, data modeling analysis, and portrait structure presentation. The data information collection in this article is mainly crawled from the Weibo platform through crawler technology. After crawling, it is divided into data and text types. Due to inconsistent data formats and duplicate values in data, different types of information need to be preprocessed separately; The user dimension selection needs to be set according to the purpose of the portrait. In order to facilitate the development of public opinion guidance, this article selects three portrait dimensions: basic attributes, comment positions and comment topics; data modeling analysis is to establish portrait models for the above three dimensions; portrait structure Presentation is to show the result of the portrait in an intuitive way for analysis and application.

The above four links are processed in chronological order to obtain a flow chart of the subject of public opinion. As shown in Figure.1, the entire process is divided into five steps: data acquisition, data type division, data preprocessing, establishment of a profile model and visualization of profile results. Among them, data acquisition and data type division uses data crawler technology to crawl the basic information, comment content and other types of data of Weibo users on Sina Weibo to form the data source of the user portrait; the portrait data needs to be previewed before starting the portrait. The processing operations mainly include missing value processing, data deduplication, and standardized processing of numerical data; then a portrait model is established from the three dimensions of basic attributes, Weibo comment stance and Weibo comment topic, and finally the three-dimensional
portrait result Fusion and display using visualization technology. The following focuses on the principles and methods of the three profile dimensions: basic attributes, Weibo comment position and Weibo comment theme.

2.1 Basic attribute dimensions
The basic attribute information of Weibo users includes nickname, gender, signature, number of followers, number of fans, comment time, etc. Among them, nickname, gender, signature and comment time can be directly used as portrait tags; the number of followers, number of fans, etc., to a certain extent reflecting the popularity of users, this article draws on the "Fame Index" (RI) [12] dimension method to quantify user reputation (see formula 1). GZ represents the number of user followers, FS represents the number of users' fans, and N represents the total number of research samples, that is, the total number of Weibo users who participated in the event comments.

\[ RI = \frac{FS}{GZ} + \frac{FS}{N} \]  

(1)

2.2 Dimensions of Weibo comments
Sentiment analysis is an indicator that many scholars often use when conducting research on Weibo comment tendency, and it can well reflect the sentiment status of Weibo users participating in comments. However, sometimes the results of sentiment analysis are used to manage public opinion, and the effect is often not ideal. For example, for negative hot events, the core opinions of Weibo users with the same sentiment category may be completely opposite, that is, they stand in completely different places, or even It is opposite.

Therefore, this article selects the position dimension of Weibo comment content for analysis. Weibo comment position is an important indicator of Weibo users. It can be derived from the comments made by Weibo users on event information. In this article, comment positions are divided into positive and negative Three types of neutral and neutral. A positive stance refers to a view that expresses sympathy, regret, or a positive attitude towards the occurrence of the "Notre Dame de Paris" incident, is consistent with the facts, and has a positive significance in guiding the public opinion of the incident, such as "I am sorry to see it as a Chinese", "Distressed, regretful, a cultural treasure of mankind! It is so precious but...so sad"; the negative position refers to the indifference and gloating attitude towards the occurrence of the "Notre Dame de Paris" incident, which is inconsistent with the facts and has a negative significance in guiding the public opinion of the incident "I’m not so great, I just feel that I deserve it." “They even burned the Old Summer Palace in front of them, just to taste the taste of their civilization and monuments being burned.” The neutral stance refers to the fact that the “Notre Dame de Paris” incident happened. Express an obvious attitude that is consistent with the facts,
but has no positive or negative significance for the public opinion guidance of the event, such as "Quasimodo lost his beloved girl, and eventually lost his beloved bell tower", "A fire will be burned together such".

This paper uses machine learning algorithms to establish an automatic classification model for Weibo user positions. First, the Weibo users’ comment positions are manually labeled, and the Weibo comment data is divided into training and test sets, and then the labeled training data is used to train the learned classification model is trained, and the model is tested with the test set data. When the accuracy of the model reaches the expected accuracy, the commentary positions of other users can be classified and marked with position labels.

2.2.1 Data set labeling and classification model preparation
The classification model based on the machine learning algorithm requires a labeled data set. Therefore, before the model is established, the user comment information needs to be commented and marked with positive, negative, and neutral tags. Positive means holding a positive view of the speech and behavior in the incident, negative means holding a negative view of the speech and behavior in the incident, and neutral means holding a neutral attitude to the speech and behavior in the incident.

The content of Weibo comments is text information, and text preprocessing is required before entering the machine learning model. The process is mainly divided into three steps. The first step is text segmentation, that is, the text comment sentence is divided into several words. This article is based on the Python programming environment and uses the Chinese word segmentation component jieba to segment the comment text. The second step is to remove stop words, that is, remove a large number of stop words in the user’s comment information, such as punctuation marks, single Chinese characters that are used very frequently, etc., mainly by calling common stop words in the Chinese stop word list carry out. The third step is text vectorization, that is, the comment text is expressed as a series of vectors that can express the semantics of the text. In this article, the text information is mainly processed through the word2vec bag of words model.

2.2.2 Oversampling processing of unbalanced data sets
By categorizing the position of the comment data with position labels, analyzing the distribution situation, it is found that there is an imbalance in the label category, and the samples with the negative category are obviously more than the samples with the positive and neutral categories. Directly using unbalanced data to train machine learning models, because the two types of positive and neutral are relatively small, the classification effect of these two types is not ideal.

Therefore, the data set needs to be oversampled before model training, so that the distribution of the number of samples of different types is basically the same. This article uses the RandomOverSampler module of the imblearn Python library to oversample the data set, and then train the model after the different types of data in the training set are evenly distributed.

2.2.3 Use machine learning algorithms to build a classification model of commentary positions
(1) Classification algorithm selection
There are many classification algorithms in commonly used machine learning algorithms, such as decision trees, K nearest neighbors, support vector machines, and naive Bayes. Among them, decision tree, K nearest neighbor and support vector machine directly calculate the mapping relationship between sample output label Y and sample feature X through learning; the principle of the naive Bayes algorithm is to calculate the joint distribution P of sample output label Y and sample feature X (X, Y), and then calculated using formula 2. Naive Bayes algorithm has the characteristics of intuition and small amount of calculation, and is suitable for the classification of a large amount of vectorized text information. Therefore, this paper uses the naive Bayes algorithm to establish a comment position classification model.

\[
P(Y \mid X) = \frac{P(X, Y)}{P(X)} \tag{2}
\]
(2) Performance evaluation of classification model

Due to the imbalance of categories in the comment data set, when evaluating model performance, in addition to the accuracy rate, the AUC value and ROC curve, which are indicators of comprehensive measurement accuracy and recall rate, should also be selected for comprehensive evaluation of model performance.

2.3 Comment topic dimensions

2.3.1 User reviews word cloud

"Word cloud" refers to a "keyword cloud" generated to highlight high-frequency "keywords" in text information, so as to more intuitively obtain the subject information of the text. In the Python programming environment, the word cloud can be generated by loading the word cloud module wordcloud, then processing the text word segmentation, removing stop words, etc., and finally by calling the generate function in the word cloud module.

2.3.2 Use LDA theme analysis

LDA (Latent Dirichlet Allocation) is a document topic generation model. The establishment process is divided into five steps: the first step is to use the word segmentation tool to segment all the comment texts to obtain the phrase sequence; the second step is to assign an ID to each word through corpora.Dictionary; the third step is to sort out each word Word frequency and form a sparse vector; the fourth step is to use the LDA model for training, and divide all review texts according to the specified number of topics. The fifth step is topic inference, giving a specific comment, judging which topic it is, and obtaining the user's topic tag.

3. Empirical analysis

3.1 Data source and acquisition

Through the web crawler technology, we can crawl the relevant data of Weibo public opinion subjects during the period from April 19 to April 23, 2018 of the "Notre Dame de Paris" incident. It mainly includes two types of data. One is the basic information of Weibo users, including gender, Age, number of fans, number of likes, number of Weibo reposts, etc.; the second is the comment text of the popular Weibo of the "Notre Dame de Paris" incident, which mainly reflects the users' own views and positions. These data form the data source of the main image of the Internet public opinion based on the big data of public opinion.

3.2 Results calculation and analysis

3.2.1 Basic attribute portrait results

The portrait results of the basic attributes are shown in Table 1. Nicknames and signatures can reflect the personalities and values of Weibo users to a certain extent; gender can be analyzed comprehensively with the following comment positions and theme portrait dimensions to explore the relevance of comment positions and topics and gender; comment time is an important time dimension The information can reflect the stage of the event when the user’s comments are published, and can provide an important reference for public opinion governance from the time dimension; the larger the reputation index, the greater its influence may be, which can be determined among the many Weibo users participating in the public opinion Governance strategy provides reference.

| Nickname | Gender | Signature | Reputation | time   |
|----------|--------|-----------|------------|--------|
| King King| male   | The people have faith, the country has strength, and the nation has hope | 44434.12 | 2018.8.21 |
3.2.2 Weibo commentary position portrait results
This article is to train a naive Bayes classification model, and manually mark the position of the comment text of the crawled "Notre Dame de Paris" incident. In the marked 480 pieces of data, the positions are "positive", "negative" and "neutral". The ratio of is 291:160:69, which shows that the data set is unbalanced. In order to improve the accuracy of the classification model, the RandomOverSampler module of the imblearn Python library was first used to oversampling the data set, and then the multi-class naive Bayes algorithm was used for classification. The final AUC value reached 95.2%, which is an improvement compared with no oversampling. The ROC curve of the classification model with and without oversampling processing is shown in Figure.2. It can be seen that the AUC value after sampling has been significantly improved, indicating the model performance after training on the sampled data set has been significantly improved.

The trained model is used to classify the 11744 comment positions crawled. Among them, the positions of "positive", "negative" and "neutral" are 5163, 4981, and 1600 respectively. Weibo users with a "positive" position and Weibo users with a "negative" position are basically 1:1, and there are relatively few "neutral" positions. Facing the same incident, there are two camps with opposing views, and effective measures need to be taken in a timely manner to carry out targeted public opinion governance.
3.2.3 Weibo comment topic portrait results

According to the LDA topic analysis steps described above, set the number of comment content topics to 5, and each topic is described in 5 words. The result is:

(0, 0.042*"No" + 0.029*"In time" + 0.010*"I have been" + 0.009*"Disappeared" + 0.009*"Paris")
(1, 0.054*"Unfortunate" + 0.042*"No" + 0.030*"I have been" + 0.029*"Good" + 0.024*"Really")
(2, 0.027*"Yuanmingyuan" + 0.026*"heartache" + 0.018*"Notre Dame de Paris" + 0.016*"burning" + 0.015*"good")
(3, 0.062*"Old Summer Palace" + 0.018*"Burn" + 0.017*"God" + 0.017*"Notre Dame de Paris" + 0.015*"Sorry")
(4, 0.049*"No" + 0.047*"Cassie" + 0.044*"Mode" + 0.043*"Beloved" + 0.030*"Bell Tower")

By analyzing the top 5 keywords of the 5 topics and referring to the review text of each topic, the core views of the 5 topics are obtained as shown in Table.2. Topics 1 and 2 both express the burning of Notre Dame de Paris. Unfortunately, the difference is theme 1 because I haven't had time to read it, and theme 2 is standing on the height of human civilization. The comments on Theme 3 and Theme 4 all think of the burning of the Old Summer Palace in our country. Theme 3 expresses "gloriousness" for the burning of Notre Dame de Paris, and theme 4 expresses regret. Theme 5 borrowed a famous saying "Quasimodo lost his beloved girl and eventually lost his beloved clock tower" to express his neutral attitude towards Notre Dame de Paris being burned.

Table 2. Weibo user comment topics

| Subject number | Subject content |
|----------------|-----------------|
| Theme 1        | I haven't had time to see Notre Dame de Paris, and I regret that it was burnt down. |
| Theme 2        | Expressed regret about the burning of Notre Dame de Paris, and considered it a loss of human civilization. |
| Theme 3        | Recalling the painful history of the burning of the Old Summer Palace in our country, I think Notre Dame de Paris burned well. |
| Theme 4        | Reminded of the burning history of Old Summer Palace, and turned sympathy into regret. |
| Theme 5        | Quasimodo lost his beloved girl and eventually lost his beloved clock tower. |

In order to obtain the subject category of each comment, it is necessary to infer the subject of each comment. According to the content of a user's comment, "Really deserves it, God is reincarnation, who is forgiven by the heavens. They were not very happy when the Yuanmingyuan was burned. Anyway, I I won't regret anything, I can only admire one sentence: 'It’s good to burn" for topic inference, for example, use lda.get_document_topics to get the output: [(0, 0.014644829), (1,
0.0147395665),(2,0.7232139),(3,0.23296784),(4,0.0144338235)] respectively represent the probability that the comment belongs to 5 topics. More available, this comment belongs to the third topic.

3.2.4 Presentation of overall portrait results
(1) Overall portrait result
Integrating the portrait information of the three dimensions of Weibo user's basic attributes, comment stance, and comment topic, the overall portrait result of the main body portrait of online public opinion based on public opinion big data is obtained, as shown in Figure.3. The overall portrait result can comprehensively and systematically describe the characteristics of each dimension of Weibo users in the event. Since the commentary position is an important reference for the management of Internet public opinion, this article takes the commentary position as the core dimension of the subject portrait of public opinion on Weibo. Commentary stance and other portrait dimensions are not isolated from each other. With commentary stance as the core, it can be integrated with dimensions such as comment time, comment topic, and reputation index for analysis and analysis.

Fig.3 Effect picture of the main portrait of online public opinion based on public opinion big data

(2) Integration of portrait results—Analysis of the reputation index of each commentary's position
The reputation index can reflect the influence of Weibo users in the event to a certain extent. In order to carry out targeted public opinion governance, it is crucial to understand the reputation index of Weibo users in various commentary positions. From Table.3, we can get a basic understanding of the reputation index of the Weibo user groups for each commentary position. It can be seen that the positions of the average reputation index from high to low are "neutral", "positive" and "negative"; "neutral" and "positive" The maximum fame index and the third quartile of the position are also higher than those of the "negative" position. It can be seen that although the number of Weibo users with "negative" positions is basically the same as that of "positive" positions, the reputation index of Weibo users is obviously lower. In the “negative” position, there is only one Weibo user with a reputation index above 10,000, a total of 7 people between 1,000-10,000, a total of 35 people between 100-1,000, a total of 236 people between 10-100, and a total of 4702 people under 10. It can be seen that there are fewer users with high reputation index, and most users have low reputation index and low influence.

| Comment topic | Positive | Negative | Neutral |
|---------------|----------|----------|---------|
| Theme 3       |          |          |         |
| Nickname: King King | 0.027"Yuanmingyuan" + 0.026"Heartache" + 0.018"Notre Dame de Paris" + 0.016"Burn" + 0.015"good" |
| Basic Attributes |          |          |         |
| Sex: Male      |          |          |         |
| Signature: The people have faith, The country has power, The nation has hope |          |          |         |
| Fame index: 74.25 |          |          |         |
| Review time: 2019.4.26 |          |          |         |

Table.3 The reputation index of Weibo user groups by commentary

|          | Positive | Negative | Neutral |
|----------|----------|----------|---------|
| Quantity | 5163     | 4981     | 1600    |
| average value | 34.24635 | 10.66648 | 81.4985 |
| Sample standard deviation | 442.9383 | 192.0216 | 2630.124 |
| Minimum | 0        | 0        | 0       |
| First quartile | 0.400896 | 0.344167 | 0.386885 |
Second quartile  0.858833  0.74058  0.864147  
Third quartile  2.046947  1.775703  2.024813  
Max  16714.41  11939.02  110160.4

(3) Integration of portrait results—changes in the time dimension of each commentary’s position

By counting the time information of each commentary position, it is known that most of the comments were published on the second day after Notre Dame de Paris was burned (April 26), and the total changes of each commentary position were counted by the hour to obtain each commentary position The number changes over time, as shown in Figure.4. It is easy to see from the figure that the three kinds of commentary positions increased sharply from 7 am to 8 am, and the growth rate from high to low was "positive", "neutral" and "negative", and the three positions from 8 am to 13:00. The growth rate slowed down, and the total number of commentary positions reached a plateau after 13:00. From the perspective of the time dimension of public opinion governance, the period of sudden increase from 7 am to 8 am is the best guiding period.

3.2.5. Public opinion governance strategy

The above shows the results of online public opinion subject portraits based on public opinion big data. The purpose of developing public opinion subject portraits is to provide accurate guidance for the development of public opinion governance, comprehensively consider the results of various portrait dimensions, and use Weibo user influence, commentary positions, and comments Portrait information such as subject, comment time, etc., propose public opinion governance strategies for different scenarios:

(1) When determining key subjects, choose a combination strategy of commentary stance and reputation index, and focus on guiding users whose commentary stance is "negative" and have a high reputation index. The Naive Bayes-based classification model proposed in this paper can well classify Weibo users’ “comment positions” based on the content of comments, and then understand the distribution of people with different positions, especially the proportion of people with “negative” positions. In public opinion governance, focus on guiding users whose comment positions are "negative" and with a high user reputation index. The influence and dissemination power of people with high “famous index” are relatively large. Therefore, in order to improve governance effects and increase the input-output ratio, priority can be given to people with high “famous index” to carry out guidance, so as to use their influence and accelerate facts The spread of truth and positive energy.

(2) When determining the time for governance of public opinion, select a combination strategy of comment position and comment time, and select the early period when the number of "negative" comment positions exceeds the threshold. The correct guidance time plays a very important role in the governance of public opinion. According to the results presented in Figure.4, the changes of the commentary position over time can be grasped in time. When the growth rate of each commentary...
position, especially the "negative" position, is the largest, it is indicated The "negative" position and public opinion began to ferment rapidly, and timely guidance at this time can achieve better results.

(3) When publishing public opinion governance content, choose a combination strategy of comment stance and comment theme, and release correct information for guidance on the theme of the "negative" comment position that is contrary to the facts. When the key subjects and the correct guidance time are determined, how to accurately push public opinion governance information becomes particularly critical. By focusing on the analysis of "negative" commentary themes, combining history and facts, and contacting professionals in relevant departments to draw up corresponding information for different topics for positive guidance, it can effectively control the growth trend of "negative" positions and grasp the trend of public opinion Development initiative.

4. Summary
As an important carrier of online public opinion expression, online public opinion has presented a big data environment. Using public opinion big data to carry out multi-dimensional portraits of online public opinion subjects can comprehensively, quickly and accurately grasp the situation of public opinion and provide effective decision support for network social governance. Based on public opinion big data and machine learning algorithms, this paper explores the problem of online public opinion subject portraits facing network social governance. From the three dimensions of Weibo users' basic attributes, Weibo comment stance, and Weibo comment theme, the fusion portrait is carried out, and the “Notre Dame de Paris” incident is taken as an example for empirical analysis. In the basic attribute dimension, the reputation index of each user is calculated using information such as the number of fans and the number of followers; in the Weibo comment position dimension, machine learning algorithms are used to classify the comment position, and the data set is oversampled using oversampling technology Processing to improve the classification effect; in the dimension of the Weibo comment topic, the LDA model is used to divide the comment data; then, based on the above portrait dimension, the fusion portrait is carried out to achieve a multi-dimensional accurate description of the main body of online public opinion. Finally, corresponding governance strategies are proposed for different scenarios.

This paper makes useful explorations in the following aspects: First, based on the big data of public opinion, we propose a framework for the integration of online public opinion subjects, that is, to integrate and portray the characteristics of public opinion from three dimensions: basic attributes, Weibo comment positions and Weibo comment topics. A panoramic depth portrait; second, in order to improve the accuracy of the Weibo comment position portrait and the overall portrait effect, the data is oversampled in the Weibo comment position dimension, which solves the problem of uneven distribution of user position labels; third is based on each dimension portrait As a result, specific portrait results fusion methods and public opinion guidance strategies are proposed for different public opinion guidance scenarios. However, due to space limitations, this article still has several shortcomings: for example, the quantification of Weibo user influence still needs to be further optimized to achieve an accurate characterization of user influence; the combination of different dimensions and weights in the fusion portrait are still In-depth research is needed to improve the effect of fusion portraits and provide more targeted and time-effective measures for the development of public opinion guidance.

Acknowledgments
This work was financially supported by the Hebei Provincial Social Science Foundation Project "Research on Internet Ideology Risk Perception and Governance Mechanism Facing Public Opinion Big Data" (Project Number: HB18ZZ009).

References
[1] Gao Guangshang. A review of research on user portrait construction methods[J]. Data Analysis and Knowledge Discovery, 2019, 3(03): 25-35.
[2] Qiu Yunfei, Zhang Weizhu. Research on group portrait construction method based on network structure and text content[J]. Library and Information Service, 2019, 63(22): 21-30.

[3] Lin Yanxia, Xie Xiangsheng. Microblog group user portrait based on social identity theory[J]. Information Theory and Practice, 2018, 41(03): 142-148.

[4] Zhao Shuguang. Social media user portraits with high conversion rate: a research based on in-depth interviews with 500 users [J]. Modern Communication (Journal of Communication University of China), 2014, 36(06): 115-120.

[5] Wei Mingzhu, Zhang Haitao, Liu Yashu, etc. Research on the Portrait of High-Influential People in Social Media Based on Multidimensional Attribute Fusion[J]. Library, Information, and Knowledge, 2019(05): 73-79+100.

[6] Ye Guanghui, Zeng Jieyan, Hu Jinglan, etc. Social Public Emotional Evolution from the Perspective of Urban Portraits[J]. Data Analysis and Knowledge Discovery, 2020, 4(04): 15-26.

[7] Ren Zhongjie, Zhang Peng, Lan Yuexin, etc. Emotional analysis of network user portraits for emergencies——Taking Tianjin "8•12" accident as an example[J]. Intelligence Magazine, 2019, 38(11): 126 -133.

[8] Deng Xiaoyi. Analysis of the impact of Weibo use on the emotional cognition of "post-95" users-from group portraits to youth culture[J]. News lovers, 2019(03): 25-28.

[9] Fan Zhe. Phase analysis of the willingness to adopt digital natives’ social media based on user portraits[J]. Modern Information, 2017, 37(06): 99-106.

[10] An Lu, Zhou Yiwen. Portraits and comparison of microblog information and comment users in the context of terrorist incidents[J]. Information Science, 2020, 38(04): 9-16.

[11] Liu Haiou, Sun Jingjing, Zhang Yaming, etc. User portraits and information dissemination behavior in online social activities[J]. Information Science, 2018, 36(12): 17-21.

[12] Wang Qiangbing, Zhang Chengzhi. Construction and application of social network user interest portraits based on gesture behavior[J]. Books and Information, 2019(02): 114-119+132.