

User Information Collection of Weibo Network Public Opinion under Python

Changhua Liu^{1,*} and Yanlin Han²

Abstract

Although the network environment is gradually improving, the virtual nature of the network is still the same fact, which has brought a great influence on the supervision of Weibo network public opinion dissemination. In order to reduce this influence, the user information of Weibo network public opinion dissemination is studied by using Python technology. Specifically, the 2019 “Ethiopian air crash” event was taken as the research subject, the relevant data were collected by using Python technology, and the data from March 10, 2019 to June 20, 2019 were constructed by using the implicit Dirichlet distribution topic model and the naive Bayes classifier. The Weibo network public opinion user identity graph model under the “Ethiopian air crash” on June 20 found that the public opinion users of ordinary netizens accounted for the highest proportion and were easily influenced by media public opinion users. This influence is not limited to ordinary netizens. Public opinion users have an influence on other types of public opinion users. That is to say, in the network public opinion space of the “Ethiopian air crash,” media public opinion users play an important role in the dissemination of network public opinion information. This research can lay a foundation for the classification and identification of user identity information types under different public opinion life cycles. Future research can start from the supervision of public opinion and the type of user identity to improve the scientific management and control of user information dissemination through Weibo network public opinion.

Keywords

Public Opinion Users, Python Technology, Weibo Network Public Opinion

1. Introduction

The emergence of social networks has had a subversive effect on the traditional public opinion dissemination mode dominated by linear communication. Compared with traditional media, social network has become the most popular information release platform and acquisition channel due to its advantages of timeliness and interactivity, which improves the communication among public opinion users in different regions, enhances the sense of community of interest, and drives the development of social media. The degree of attention and participation of non-stakeholders in emergencies [1]. With the continuous deepening and development of Internet applications, the speed and influence of network public opinion are constantly increasing, and the scope of influence is also expanding. The life cycle of online public opinion will advance in a relatively short period of time. The number and energy of discourses of social network public opinion users are constantly increasing, as well as their discourse

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received April 29, 2022; first revision August 23, 2022; second revision November 22, 2022; accepted December 11, 2022.

*Corresponding Author: Changhua Liu (chiou08723411@163.com)

¹ Research Institute of Journalism & Communication, Communication University of Zhejiang, Zhejiang, China (chiou08723411@163.com)

² Dept. of Basic Education, Boya Whole-Person Education College, Jiangsu Vocational Institute of Commerce, Nanjing, China (hanyanlin81@yeah.net)

awareness and ability. Users share and exchange information with the help of social networks, and continuously expand the radius of information dissemination [2]. Among social network public opinion users, there are active public opinion leaders who can influence their associated users, and they are “active elements” that exert influence on network public opinion users. These key public opinion nodes have become the vane of mainstream opinion expression, have strong social influence and public opinion guiding ability, and play a decisive role in the dissemination of network public opinion [3]. The study of social network public opinion user relationship includes the identification of opinion leaders, the influence strength of nodes, the classification of stakeholders, the analysis of user topic dissemination paths. It is a key element to effectively guide the direction of network public opinion and shape mainstream public opinion [4]. To sum up, it can be seen that the information exchanged and the attitude conveyed by social network public opinion users may become the key factors that interfere with the trend of network public opinion. It is an important part of network public opinion user management and information dissemination to construct the network public opinion user theme map and comprehensively analyze the identity information of network public opinion users and the relationship between the user community. However, at present, the difficult problem of supervision of online public opinion dissemination has not been well solved. Based on this, this paper will use Python technology to study the user information of Weibo network public opinion dissemination, in order to better realize the management of network public opinion users and the dissemination of public opinion information. The structure of the paper is as follows. Section 2 analyzes the user identity identification model of Weibo network public opinion communication, including the hidden Dirichlet distribution model, identity characteristics and the construction and generative model of Weibo network public opinion communication characteristics. Section 3 constructs the user identity map model of public opinion communication on the Weibo network. Section 4 describes the systematic research design. Section 5 presents the data results and Section 6 is a summary.

2. User Identification Model of Microblog Network Public Opinion Dissemination

2.1 Endirichlet Distribution Model and Its Process Overview

The latent Dirichlet allocation (LDA) model is a topic model, which is not only used in discrete data processing (document set and word bag model), but also in natural language understanding and information retrieval [5,6]. By applying the LDA topic model to semantic mining, the representation dimension of text can be reduced, and the deep semantic features of a text can be extracted. The concepts involved in this topic model can be divided into three categories, including word, document, and topic. A word refers to a discrete unit, and a document refers to a data object to be processed. A bag-of-words consists of a group of words that indicate the order in which words can be ignored in a document. Topic models do not exclude data modeled in bag-of-words. The topic here refers to the probability distribution of the whole group of words. As a representative of the social network platform, Weibo is defined as follows: “word” refers to Chinese words with independent meaning, i.e., the result of Chinese word segmentation, “document” refers to a piece of microblog information, and “topic” refers to the topic concept concerned by public opinion spread by users in the public opinion space of the microblog

network. The sampling process of the LDA topic model is as follows. First, the topic distribution of microblog information is sampled in the Dirichlet distribution.

2.2 Construction of Identity Characteristics and Microblog Network Public Opinion Communication Characteristics

In order to ensure the normal implementation of public opinion control, most microblog platforms will include an identity information discrimination dimension, so that the algorithm can realize the identity screening mechanism, verify the authenticity of user identity information, and reduce the interference of malicious users (network water army, etc.) by means of traffic restriction. According to the survey, the number of fans is usually far less than the number of malicious users, so the authenticity of user identity information can be verified by comparing the difference between the number of fans and the number of followers. There is a certain internal relationship between the activity level of online public opinion users and the number of microblogs they have posted. When the activity level is high, the user's identity information can be considered authentic and reliable. At the same time, the identity authentication measures provided by the Weibo platform can also be used as an effective means of identity information recognition. The definition of user identity characteristics is:

$$\text{Identity} = \text{verify} + \log(e^{\text{follower}-\text{following}} + e^{\text{num}}). \quad (1)$$

In Eq. (1), whether user A passes the authentication mechanism of the microblog platform is represented by verify, the number of fans is represented by follower, the number of followers is represented by following, and the number of tweets is represented by num. Before calculating user identity characteristics, z-score normalization is also required:

$$Z = \frac{x-\mu}{\sigma}. \quad (2)$$

In Eq. (2), the characteristic value to be standardized is represented by x , and the mean value and standard deviation are represented by μ and σ , respectively. As for the construction of the characteristics of microblog network public opinion communication, the characteristics of public opinion information communication determine the influence of public opinion information, which has three specific manifestations, including the number of comments, the number forwards, and the number of likes. In addition, these three forms of expression are the decisive factors of the ability of public opinion information dissemination. Therefore, it is decided to take the number of users' own fans, the number of comments, the number of forwards, and the number of likes on public opinion information as the basic parameter values, and the set propagation characteristics are shown below:

$$\text{Propogation} = \log(e^{\text{follower}} + e^{\text{repost}} + e^{\text{comment}}) + \text{like}. \quad (3)$$

In Eq. (3), the number of fans is represented by follower, the number of forwards is represented by repost, the number of comments is represented by comment, and the number of likes is represented by like. The value obtained after z-score normalization is the prorogation value. The larger the value is, the stronger the communication ability of online public opinions will be. There is a positive correlation between them.

2.3 Naive Bayes Classifier

Naive Bayes classifier is a generative model that can estimate the posterior probability by the prior probability [7,8]. Compared with other machine learning classifiers, naive Bayes classifier has stronger parsing, simpler classification process and faster classification speed. Therefore, this paper applies it to the classification of users of Weibo network public opinion dissemination. Specifically, assuming that microblog message A is represented by D, A can be abstracted into D word bag model, A document composed of nw words, which is defined as: $D = \{w_1, w_2, \dots, w_n\}$. Let the category of $C = \{c_1, c_2, \dots, c_2\}$ microblog information release user be T, then the user identity classification can be calculated by:

$$C_{NB} = \arg \max_{c_j \in C} P(c_j) \prod_{i=1}^n P(w_i | c_j). \quad (4)$$

In Eq. (4), the prior probability of the user category is represented by $P(c_i)$, and the probability calculation of the public opinion information can be realized according to the disclosed user identity information. When the user identity category is c_j , the conditional probability of its occurrence is represented by $P(w_i | c_j)$. Based on the idea of supervised learning, the data of Weibo users are collected after the identity authentication is completed, and the prior probability $P(c_j)$ is calculated by:

$$P(c_j) = \frac{Doc(c_j)}{\sum_{c_j \in C} Doc(c_j)}. \quad (5)$$

In Eq. (5), the number of user identity category c_j is represented by $Doc(c_j)$. The posterior probability $P(w_i | c_j)$ can be calculated by:

$$P(w_i, c_j) = \frac{Weight(w_i, c_j)}{\sum_{i=1}^n Weight(w_i, c_j)}. \quad (6)$$

In Eq. (6), the weight of word w_i in user category is represented by $Weight(w_i, c_j)$, and the weight sum of all words in user identity category c_j is represented by $\sum_{i=1}^n Weight(w_i, c_j)$. There are three performance indicators for evaluation classifier. The first indicator is precision, which refers to the ratio between the number of texts correctly classified into a certain category and the number of Chinese texts in all categories, as shown in Eq. (7). The second indicator is recall rate, which refers to the ratio between the number of texts correctly classified into a certain category and the number of texts correctly classified into a certain category, as shown in Eq. (8). The range of both values is [0,1]. The closer the two values are to 1, the higher the precision and recall rate will be. The third indicator is F_1 value, which can be regarded as the weighted average of precision and recall rate. The calculation formulas of the three indicators are:

$$Precision = \frac{\sum_{c_i \in C} True(c_i)}{\sum_{c_i \in C} Doc(c_i)}, \quad (7)$$

$$Recall = \frac{\sum_{c_i \in C} True(c_i)}{\sum_{c_i \in C} Response(c_i)}, \quad (8)$$

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \times 100\%. \quad (9)$$

3. Construction of User Identity Atlas Model for Weibo Network Public Opinion Dissemination

3.1 Construction Idea

In the microblog network public opinion space, the opinions of public opinion users can be used as reference factors for others to make decisions, and they can be regarded as stakeholders to some extent. Network public opinion users usually play different information roles during their participation in public opinion dissemination. The user identity map of Weibo network public opinion dissemination is modeled in the form of a graph, and the identity of network public opinion users is divided according to the different topics they pay attention to. Specifically, the network public opinion users are regarded as nodes and the forwarding and commenting relationships are considered as edges, and then the identity of public opinion users is visualized in a specific public opinion network space. After completing the construction of user identity pictures, the classification of public opinion users can be intuitively understood, and the participation of public opinion users in different public opinion life cycles can be understood, which is conducive to the reasonable control of public opinion supervision departments.

3.2 User Identity Graph Model Construction based on Endirichlet Distribution and Naive Bayes

The construction process of the user identity map model based on LDA and Naive Bayes is as follows. First, the Python web crawler [9,10] is used to obtain microblog forwarding comment data, and the Jieba word segmentation tool is used for word segmentation, and the stop word operation is used for preprocessing. Secondly, using the document-topic distribution trained by the LDA topic model, the semantic characteristics of online public opinion users are deeply mined. According to the microblog

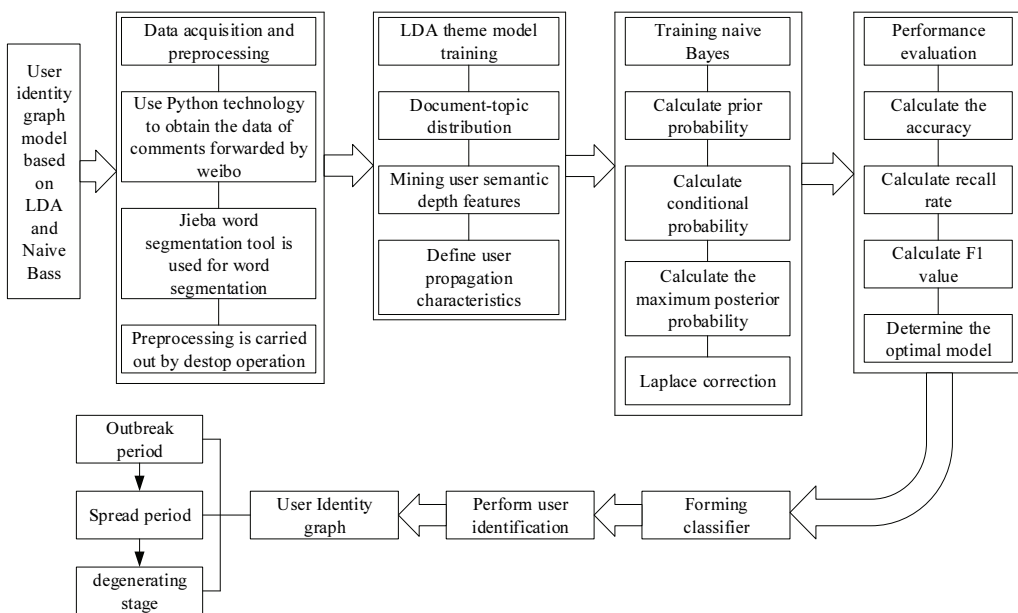


Fig. 1. Weibo network public opinion user identity atlas model.

authentication, the number of followers, the number of fans and the number of Weibo posts, the identity characteristics of users are defined. Meanwhile, the communication characteristics of users are defined based on the number of forwards, the number of fans, the number of likes and the number of comments. Then, users who have completed microblog authentication are used as supervised data to train naive Bayes. Specifically, the corresponding prior probability, conditional probability and maximum posterior estimate are obtained and Laplacian correction is performed. Finally, the F_1 value was determined by the precision and recall rate, and the result was used as the performance index of the model's hyperparameters, and the optimal model was determined by repeated testing. After the classifier is formed, it is used to realize the classification of users who have not completed Weibo authentication, check the identity information of different categories of users, obtain the user identity map under the network public opinion life cycle, and display the identity map of public opinion users at different stages of a life cycle, as shown in Fig. 1.

4. Research Design

4.1 Data Sources

As of March 2021, Weibo had 530 million monthly active users and 230 million daily active users, according to a report published by QuestMobile. With the differentiation adjustment of domestic Internet products, the whole ecological chain product line of the Weibo multi-terminal platform (PC, Android, WAP, iOS, etc.) has been realized. Compared with traditional social networking platforms, Weibo pays more attention to users' free choice of information acquisition. Users can select a user to follow according to their preferences. After the so-called "fan" relationship is established, users can constantly acquire more information content according to their preferences all the time and interact with each other. Therefore, for users with different identities, the influence of public opinion the Weibo network is relatively flexible. Generally, the more fans the user follows, the more potential energy the user follows to spread online public opinion. In the online public opinion space of the emergency "Ethiopian air crash", this paper takes the forwarded comment information as the data source of this study, and takes the users who have an intersection with this topic (comment, forward, like, etc.) as the information people, and then builds the user information atlas in the life cycle of online public opinion.

4.2 Data Collection

In data collection, this paper uses Python technology to obtain user-related data. Specifically, first of all, Baidu index is used to determine the data collection period, as shown in Fig. 2.



Fig. 2. Data collection period.

As can be seen from Fig. 2, the “Ethiopian air crash” was the starting point of the active period on March 10, 2019, and reached the maximum value in the active period on March 14, 2019, and then ended on June 20, 2019. With the help of the advanced search function in Weibo, select the start time and enter the keyword “Ethiopian air crash” to search. Secondly, uniform resource locator (URL) is built, Python technology is used to climb all web pages in a request, and the number of pages returned in this request is recorded. Then, a page ORL is built to store web page data. Then, XML Path Language (XPath) is used to parse the data obtained by crawling to obtain relevant information, such as microblog content, blogger information, blogger address and nickname, and then collect ID, personal information, and other fields. Finally, this request is ended, and it jumps to the next request to collect the information released by relevant bloggers in a circular manner. Finally, the result is that there are 21,573 users forwarding comment data by network public opinion users, and there are 34,346 forwarded comment data.

4.3 Data Processing

In the online public opinion space, in order to clarify the user information collection of public opinion participants and stakeholders, politically sensitive data is removed and noisy data such as sales activities, business promotion and evaluation, and voting, is retained. At the same time, the forwarding comment data is preprocessed. Specifically, the Jieba word segmentation tool is first used to process the forwarding comment data. Secondly, stop words are used to eliminate stop words in the document. Finally, special text (emojicons, external links, etc.) is filtered. As for the previous LDA topic model, it is usually limited by the length of the document and the number of words, so it cannot achieve the ideal result of the topic discovery. Therefore, in order to avoid such problems, text aggregation is chosen to solve them. Specifically, according to the life cycle of the Weibo network public opinion communication, it is divided into three life cycles, namely the outbreak period, the spread period and the decline period, and the time of forwarding comments is used as the basis of dimension division. Within the same online public opinion cycle, a document is edited for each user to record all the information published by them. At the same time, in order to train the naive Bayes classifier, in the network public opinion space environment, authenticated users are used as the supervision data to provide training sets and test sets for the training of the naive Bayes classifier, so as to obtain the optimal naive Bayes classifier.

4.4 Weibo Network Public Opinion Event Overview and Life Cycle Division

In order to ensure a reasonable division of the life cycle of Weibo network public opinion, the overall overview of the “Ethiopian air crash” network public opinion event is sorted out here, as follows. On March 10, 2019, the official Weibo account of CCTV news revealed that a Boeing 738MAX-8 aircraft of Ethiopian Airlines and Obia Airlines lost contact shortly after takeoff (6 minutes) and crashed near Bishoftu, about 45 km from the capital. There were 157 crew members on board, including 8 crew members and 149 passengers. None of the passengers survived, and Chinese nationals were confirmed on board. At 6:00 pm on March 11, the black box was successfully found and brought back by the Ethiopian aviation staff for testing. There are two main parts to be tested: the flight data recorder and the voice recorder. This news event means the beginning of the network public opinion event of the “Ethiopian air crash.” This sudden public opinion event quickly attracted the attention of the majority of Weibo network public opinion users, which also means that this network public opinion entered the outbreak period. The Ethiopian Prime Minister later declared March 11, as a national mourning day to

honor the victims of the incident. On March 12, a number of countries imposed commercial bans on Boeing 738MAX-8 aircraft. The United States and Canada had insisted that the plane could fly normally, but under pressure from public opinion, they issued a grounding order. On March 14, relatives of the victims arrived at the site and held a memorial service. On March 17, Ethiopia’s Transport Ministry said there were similarities between the crash and that of a Lion Air Boeing 737-8 in Indonesia. Meanwhile, on March 29, the Ethiopian Ministry of Transport released the results of its investigation into the crash, indicating that there were no abnormalities before or during the aircraft's takeoff and that the pilot did not commit any violations. To date, the Ethiopian Ministry of Transport has made two recommendations: first, that Boeing re-examine the control line of the aircraft’s operating system; and second, that regulators inspect the 737MAX8 before returning the aircraft to service after confirming there are no potential risks. After this news broke, Boeing was pushed to the commanding heights of public opinion. March 17–April 12 can be regarded as the spreading period of online public opinion. In the space environment of Weibo public opinion, all users start to interact frequently and express different views on the same topic. On April 12, a Boeing representative said that a software update for the aircraft introduced in the incident was “working fine.” This statement did not directly give the relevant content about whether there are safety risks of the aircraft, but with the passage of time, the heat of the topic gradually faded, and the online public opinion gradually shifted from “Ethiopian air crash” to “Boeing Company safety risks,” which also means that “Ethiopian air crash” officially entered a period of decline. The specific development of the Ethiopian air disaster is shown in Table 1.

Table 1. Time development overview of “Ethiopian air crash”

Time	Critical event
2019.03.10	An Ethiopian Airlines Boeing 738MAX-8 has crashed
2019.03.12	Several countries have issued flight bans on the crashed aircraft
2019.03.14	Relatives of the victims arrived to pay their respects
2019.03.17	The Ethiopian Transport Ministry said there were similarities between the technology and the Lion Air Boeing 737-8 crash in Indonesia
2019.03.29	The results of the investigation confirmed the pilot's conduct
2019.04.12	A Boeing representative spoke about the crash

According to the characteristics of online public opinion during the “Ethiopian air crash” and the division of public opinion events above, the whole cycle of online public opinion events is finally determined, i.e., it is divided into three stages, such as outbreak period, spread period and decline period, as shown in Fig. 3.

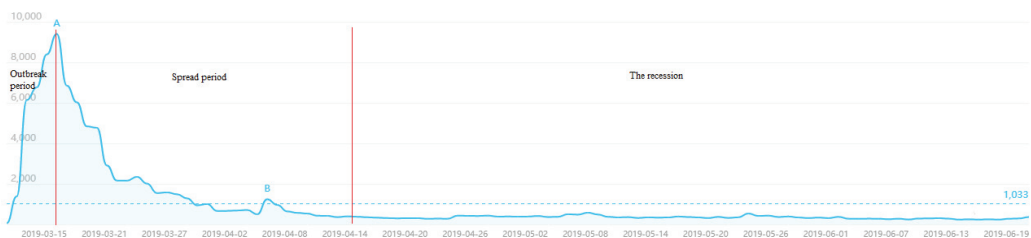


Fig. 3. Division of online public opinion event cycle of “Ethiopian air crash.”

As can be seen in Fig. 3, the topic heat of the “Ethiopian air crash” reached its maximum between March 10 and March 14, 2019, showing the characteristics of an outbreak period. The heat of the topic decreased rapidly after March 14 but increased slightly on April 12, which was the propagation period of the event. From April 13 to June 20, the topic heat no longer picked up, although there were still online public opinion users participating in the event, indicating that the online public opinion event has entered a declining period.

5. Data Results

5.1 Classification of User Concerns

In order to ensure the accuracy of the division of the subject of user attention, the value range of the number of topics is set to ensure that it is 20 more. At the same time, in order to ensure the independence of all topics, the number of topics is determined by using topic principal component analysis and confusion curve. Specifically, the first is to determine the range of the number of topics. After testing, it is reasonable to choose 24 or 27. Secondly, the number of topics is selected 24 and 27, and the document-topic matrix and subject-word matrix are calculated according to the order of selection. Finally, two-dimensional principal component analysis is performed on all topic-word matrices to obtain the corresponding feature information, as shown in Fig. 4.

In Fig. 4, the data in the ellipse represents the number of topics, and the area of the ellipse represents the number of microblog forwards and comments under the topic. The larger the area, the higher the number of forwards and comments. The distance between the centers indicates the correlation between the topics, and the larger the distance, the smaller the correlation between the topics, and vice versa. Through observation, the number of topics with a low degree of overlap is selected as the final number of topics, so the number of topics is selected as 24.

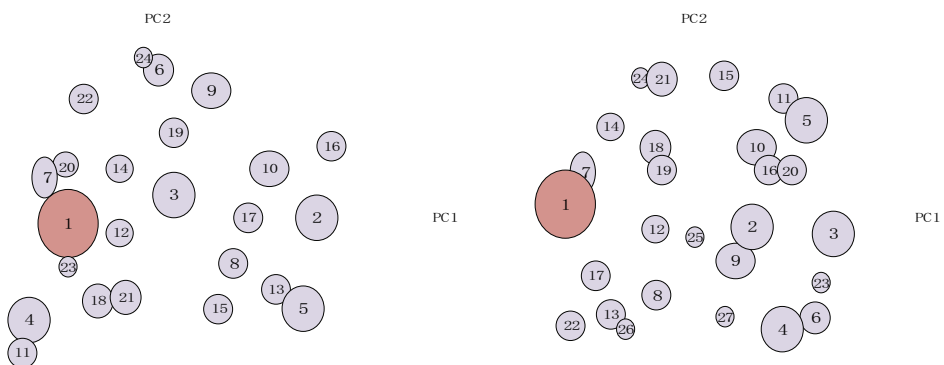


Fig. 4. Two-dimensional principal component analysis in the space of online public opinion on the “Ethiopian air crash.”

5.2 User Identification

After preprocessing the forwarded comment data, the topic text is transformed into a discrete word bag model. Based on the word bag model, the public opinion users are considered as classification labels after

the authentication is completed, and then the naive Bayes classifier is trained. Then, the value of Eq. (1) is calculated and used as an index to select the hyperparameters of the model. In the test set, the model with the best performance evaluation results is selected as the naive Bayes model. The identity authentication information of online public opinion users was selected as the label data of the training set. Among all (34,176) online public opinion users who participated in topic forwarding, a total of 6,195 users completed the identity authentication. According to the authentication information of Weibo network public opinion users, the identities of network public opinion users are classified here, including aviation industry, media, ordinary netizens and relevant enterprises. Among them, the aviation industry is divided into Ethiopian Airlines, China Airlines, and Malaysia Airlines; media is divided into government media, web media and foreign media. Ordinary netizens are divided into ordinary organizations, students, ordinary people and ordinary institutions. Relevant enterprises are divided into Boeing Company, legal consulting and psychological investigation, as shown in Fig. 5.

To ensure the classification effect of the naive Bayes classifier, the following processing steps are needed. First, the document-subject distribution of user-forwarded comment data information is obtained by means of the LDA principal model; Second, after defining the number of online public opinion users' attention, the number of fans and the number of microblog posts, the z-score is used for standardized processing, and then Eq. (1) is used to calculate the identity characteristics of online public opinion users. Thirdly, Eq. (3) is used to calculate the propagation characteristics of online public opinion users. Fourthly, 4956 pieces of data of 80% of network public opinion users who have completed identity authentication are taken as the training set, and the remaining 1,239 pieces of data are taken as the verification set. After the calculation of prior probability, conditional probability and delayed probability is completed, the naive Bayes classifier is trained. The performance index of the classifier is tested by the validation set, and the value of F_1 is 0.7896, which confirms the classification effect of the naive Bayes classifier and can train the labeled text.

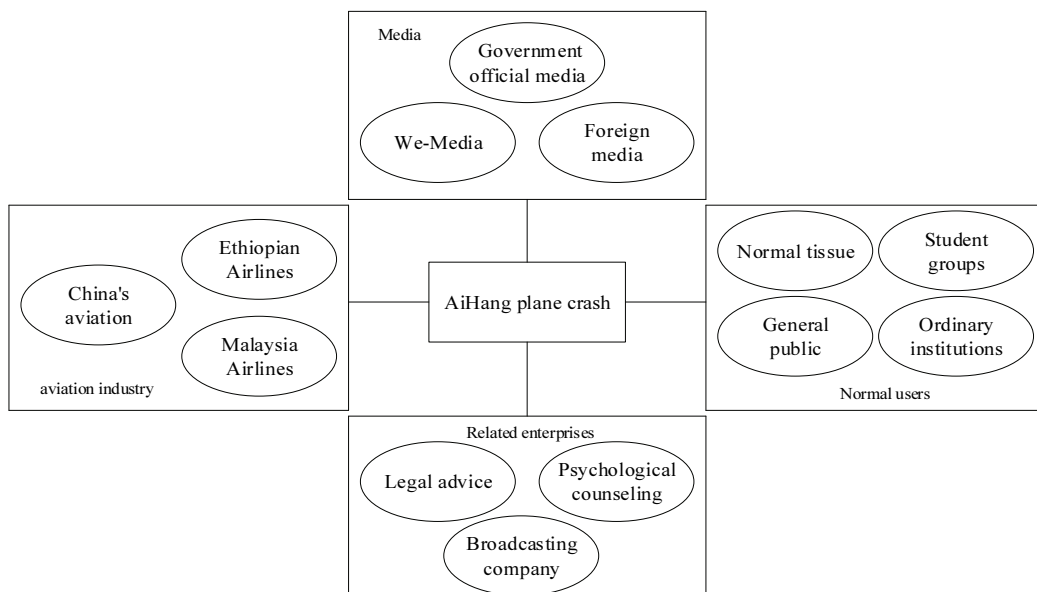


Fig. 5. User identity classification of online public opinion about the Ethiopian air crash.

5.3 User Identity Map Construction

Three hypotheses need to be made before the construction of the user identity map of the Weibo network public opinion.

Hypothesis 1: in the network public opinion space, users' forwarded comment information is regarded as the topic tendency;

Hypothesis 2: there are certain differences in comments forwarded by online public opinion users with different identity categories; and

Hypothesis 3: there is no comment information inconsistent with identity attributes among online public opinion users.

After setting the hypotheses, the online public opinion space of the “Ethiopian air crash” is divided into topics, and the online public opinion users who have not completed the authentication of identity information are identified. After the identification of users in the online public opinion space is implemented, the identity map of Weibo network users is constructed. Specifically, this paper takes “Ethiopian air crash” as the keyword and analyzes the valid microblog data (34,327) collected from March 10, 2019 to June 20, 2019. With the public opinion users of the network as the node and the comments and forwarding relationships as the edge, Neo4j is used to construct the identity map of the public opinion users of the microblog network. Some of the results are shown in Fig. 6.

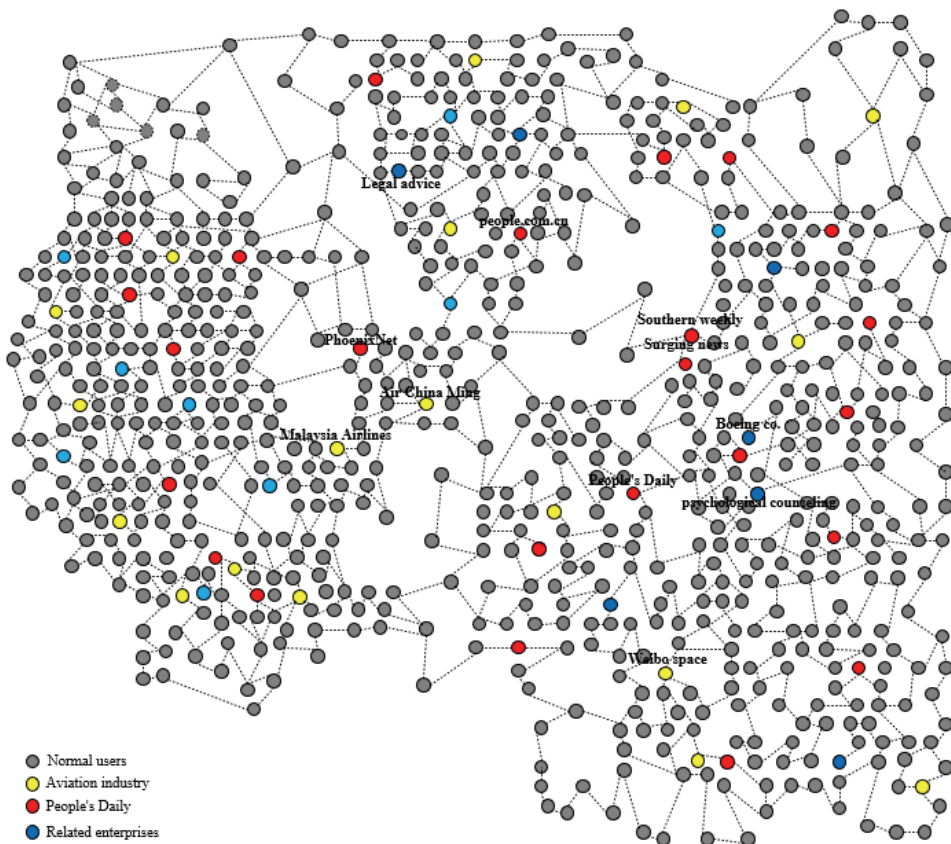


Fig. 6. User identity map of Weibo online public opinion (partial).

As can be seen in Fig. 6, due to the differences in communication characteristics and identity characteristics, information agents present different user identities. Ordinary users and related enterprise users are basically centered on aviation public opinion users and media public opinion users, and want to spread around. The radius of transmission is proportional to the coverage, and then increases and increases. Although ordinary netizens account for the largest proportion of public opinion users, these public opinion users are vulnerable to the influence of media public opinion users. This influence is not limited to ordinary netizens, but also has a certain influence on other netizens. This means that media public opinion users play an important role in the dissemination of online public opinion information in the online public opinion space of the "Ethiopian air crash." They play two roles: one is the disseminator of information; secondly, information producers have played a positive role in promoting the dissemination of online public opinion information. Ordinary netizens, enterprises and airlines have also made contributions to the development of online public opinion communication, such as reducing barriers to information communication and promoting the development of a small-world network to a large-world network.

6. Conclusion

To sum up, this paper first describes the user identity identification model of microblog network public opinion propagation, including the Endirichlet allocation model, user identity characteristics and microblog network public opinion propagation characteristics, naive Bayes classifier, which lays a foundation for the construction of user identity graph model. Secondly, based on the above theoretical research, the user identity atlas model based on LDA and naive Bayes is constructed. Finally, according to the model, an empirical study is conducted on the emergency public opinion event "Ethiopian air crash" in 2019, and the results are as follows. In terms of the classification of users' concerns, the number of subjects is determined (24). Under this number, the subjects can maintain good independence. In terms of user identification, the value of F_1 calculated is 0.7896, which confirms the classification efficiency of the naive Bayes classifier. In the construction of a user identity map, it is found that different types of public opinion disseminators have made due contributions to the spread of online public opinion information. This study can further promote the rationality and scientificity of the network management of public opinion on microblog network, and make a contribution to the improvement of the supervision of network public opinion dissemination.

References

- [1] Q. Zhao and X. Wei, "Analysis of user behavior influence in social network public opinion events: taking "Changchun Changsheng Vaccine Incident" as an example," *Journal of Hubei University of Economics (Humanities and Social Sciences Edition)*, vol. 2020, no. 11, pp. 14-17, 2020.
- [2] M. Huang and L. Yuan, "Audience involvement, comment interaction and public opinion: a case study of user comments on NetEase News based on Discourse analysis," *Future Communication*, vol. 28, no. 2, pp. 69-83, 2021.
- [3] L. Jing and H. Su, "Research on the application of government micro-blog from the perspective of holistic governance," *Frontier Economy and Culture*, vol. 2022, no. 12, pp. 1-3, 2022.

- [4] R. Tang, and C. Liu, "Integrity and differentiation of international communication in the context of great changes," *Modern Communication*, vol. 43, no. 4, pp. 75-79, 2021.
- [5] Y. Zhang, X. Han, M. Yang, M. Wang, L. Zhang, P. Ye, and B. Xu, "Distributionally robust unit commitment based on imprecise Dirichlet model," *Proceedings of the Chinese Society for Electrical Engineering*, vol. 39, no. 17, pp. 5074-5084+5288, 2019.
- [6] W. He, H. Xie, and G. Feng, "Review on latent Dirichlet allocation model," *Journal of Information Resources Management*, vol. 8, no. 1, pp. 55-64, 2018.
- [7] H. J. Cui, R. Zhao, M. Q. Zhu, and X. Li, "Travel attributes analysis of passengers based on Naive Bayes classifier," *Science Technology and Engineering*, vol. 20, no. 11, pp. 4572-4576, 2020.
- [8] G. Zhao, J. Lai, Y. Chen, H. Sun, and Y. Zhang, "Residents' travel origin and destination identification method based on naive Bayes classification," *Journal of Computer Applications*, vol. 40, no. 1, pp. 36-42, 2020.
- [9] C. Li, "Design and analysis of web data crawler based on Python," *China Computer & Communication*, vol. 32, no. 24, pp. 130-132, 2020.
- [10] P. Sun and G. Lv, "On the collection and analysis of COVID-19 data by using Python," *Journal of Hebei Institute of Architecture and Civil Engineering*, vol. 38, no. 4, pp. 155-160, 2020.



Changhua Liu <https://orcid.org/0000-0001-9805-3757>

He has received the doctor's degree. He is working in Communication University of Zhejiang, China. He is interested in digital communication and governance, discourse communication.



Yanlin Han <https://orcid.org/0009-0000-9152-072X>

He was born in April 1981, has received the master's degree from School of Mathematics and Statistics of Jiangsu Normal University in June 2006. He is a lecturer. He is interested in applied mathematics.