Analysis of Social Media Utilization based on Big Data-Focusing on the Chinese Government Weibo

Xiang Li[1], Xiaoqin Guo[2], Soo Kyun Kim[3], and Hyukku Lee[4*]
[1] School of Business, Heze University, Heze 274015, Shandong, China
[Email: lixiang@hezeu.edu.cn]
[2] College of Politics and Law, Heze University, Heze 274015, Shandong, China
[Email: guoxiaoqin@hezeu.edu.cn]
[3] Department of Computer Engineering, Jeju National University, Jeju 63243, Korea
[Email: kimsk@jejunu.ac.kr]
[4*] Department of East-Asia Studies, Graduate School, Pai Chai University, Daejeon 35345, Korea
[Email: leehk@pcu.ac.kr]
*Corresponding author: Hyukku Lee

Received June 6, 2022; revised July 6, 2022; accepted July 12, 2022; published August 31, 2022

Abstract

The rapid popularity of government social media has generated huge amounts of text data, and the analysis of these data has gradually become the focus of digital government research. This study uses Python language to analyze the big data of the Chinese provincial government Weibo. First, this study uses a web crawler approach to collect and statistically describe over 360,000 data from 31 provincial government microblogs in China, covering the period from January 2018 to April 2022. Second, a word separation engine is constructed and these text data are analyzed using word cloud word frequencies as well as semantic relationships. Finally, the text data were analyzed for sentiment using natural language processing methods, and the text topics were studied using LDA algorithm. The results of this study show that, first, the number and scale of posts on the Chinese government Weibo have grown rapidly. Second, government Weibo has certain social attributes, and the epidemics, people's livelihood, and services have become the focus of government Weibo. Third, the contents of government Weibo account for more than 30% of negative sentiments. The classified topics show that the epidemics and epidemic prevention and control overshadowed the other topics, which inhibits the diversification of government Weibo.

Keywords: Big Data, Social Media, Text Mining, LDA algorithm, Python, Government Weibo

1. Introduction

Driven by the rapid popularity of the mobile Internet, social media is one of the fast-growing digital channels. According to the Global Social Media Study 2021, the open, communicative, and engaging features and qualities of social media have attracted an increasing number of users, reaching 4.3 billion active users worldwide by 2021[1, 2]. As a result, social media is revered by governments as an important platform for service delivery and government-public interaction, and as a common digital resource for different governments and public administrations[3]. The use of social media in government is changing the landscape of public institutions around the world[4].

As one of the most popular social media tools in China, the government Weibo is a typical means for the Chinese government to implement digital governance. Compared to traditional government websites, government Weibo is more convenient, transparent, and effective in terms of service and interaction, and more interactive in terms of real-time information dissemination in case of emergencies, which has contributed to the rapid development of government Weibo. At present, all provincial-level administrative regions in China have built various types of government Weibo, which covers most areas of provincial government agencies and have become an important platform for the Chinese government to ask questions, understand public opinion, gather public wisdom, and communicate with the government and the people. In recent years, government Weibo has become an important window for government affairs disclosure, service provision, and government-civilian interaction, and also plays an important role in emergency management. Its development and use level will also profoundly affect the modernization of the national governance capacity and system. At the public level, they can realize their basic desire to participate in politics and discuss politics by retweeting and commenting on government Weibo.

The interaction between the government and the public on the Weibo platform generates a large amount of valuable information, which can be used by the government departments to grasp the public opinion and the related public opinion situation in the hot events based on Weibo. However, due to the large scale and unstructured data of these information resources, it is necessary to obtain valuable information directly from them with the help of relevant methods of big data analysis. To better understand the development status and problems of government Weibo in China at the provincial level, this study collects data related to government Weibo in 31 provinces between January 2018 and April 2022, and discusses the development status, posting content, and topics involved in government Weibo in recent years more comprehensively through textual analysis and unsupervised learning methods. Based on the LDA(Latent Dirichlet Allocation) topic model[5], this study identifies the topics of concern in the content of Weibo of each provincial government. In this way, relevant policy suggestions are put forward for the optimization and improvement of government Weibo.

This paper is organized as follows. Section 2 provides an overview of the literature, including the use and development, advantages, and values of big data and government Weibo. The third section is data and research methods, which introduces the methods of data collection and big data analysis. The fourth section presents the analysis results using data visualization in a comprehensive way, including word frequencies, word cloud analysis, semantic relationship graphs, and topic modeling results. The conclusion section summarizes this study and discusses the research contributions as well as the shortcomings.

2. Literature Review

2.1 Big Data and Government Social Media

Big data is a huge amount of information represented by numbers such as graphics, tables, text, videos, etc. These numbers will be optimized through network interconnection and computer "brain" analysis and "algorithm" to produce a large amount of effective information that can help the government or other social organizations and individuals to make decisions. The government's "digital governance" has thus become possible[6]. When users use social media platforms, they generate a lot of data and information[7], such as content posted by users, interactive connections, or geospatial data. Big data methods can help us analyze these large-scale datasets to create public value and study social behavior[8, 9]. Rogge et al.(2017) point out that big data approaches can improve public services by enabling dynamic performance monitoring[10]. Kowalski et al.(2020) argue that big data technologies are a new approach to public decision-making using social media datasets[11].

Large-scale unstructured government Weibo data collection and analysis are difficult to analyze using traditional methods. First, the dataset contains too much content and needs to be analyzed with the help of computer technology or even artificial intelligence techniques. Secondly, the obtained data may consist mainly of unstructured textual data for which traditional statistical methods are not suitable[12]. Finally, scholars have also highlighted the ethical issues inherent in processing online data, with concerns about respecting individual privacy[8]. Anastasopoulos and Whitford(2019) argue that tools developed to deal with complex social media data must aim to address these issues while adding value to public service research[13]. The mainstream processing approach is the use of natural language processing (NLP), e.g. using topic models to summarize a large number of comments to obtain actionable insights[5, 14]. Machine learning can also be a useful tool for summarizing textual comments and efficiently retrieving relevant insights[13]. By using big data from the Chinese governmental Weibo, we applied the above methodology to this study.

2.2 Government Weibo

Social media applications are widely used by government departments in areas such as information dissemination and online policymaking and have greatly improved the transparency, engagement, and collaboration of government administration[15, 16]. Government departments are increasingly using social media platforms to interact with citizens and enhance citizen participation in government policy-making[17], and the provision of quality information by government departments on social media affects public trust in government[1, 18, 19]. Criado et al. (2013) argue that the interactive communication between the public and government and other departments on social media generates a large amount of data resources that governments and decision-makers can benefit from. This data can help governments better understand public needs and attitudes and improve the accuracy of services and governance. Governments need to use social media to obtain valuable information and insights from the public on relevant issues or emergencies[4]. Stieglitz et al. (2013) argue that governments can improve their interactions with the public and their governance by continuously collecting, analyzing, summarizing, and visualizing relevant data on social media[20]. Mansoor (2021) highlights the use of social media by government agencies to enhance citizens' trust in government during COVID-19[19].

Currently, the Chinese government's use of social media is still at the information dissemination stage, mostly using social media platforms for news updates or important announcements, and the level of interaction is not high[21]. Xiao et al. (2014) explored the

role of online public opinion and government Weibo in the modernization construction of national governance and concluded that government Weibo is one of the main ways to express a public opinion online, effectively compensating for the deficiency of traditional media in reflecting public opinion and providing a new platform for public decision-making[22]. Lu et al. (2016) used social representation theory as the main theoretical perspective to study the Chinese government Weibo, arguing that government Weibo improve citizens' perceptions of government, but the government cannot achieve results such as trust and accountability using social media alone[23]. Meng and Zheng (2017) analyzed the massive data of the Chinese government Weibo to explore the role transformation of government in social media, focusing on the communication functions of social media. Convenience services are its core communication function, with additional functions such as life information, government affairs disclosure, and political performance declaration[24]. Wang and Wang (2017) used text analysis techniques to study three provincial government Weibo in Beijing, Shanghai, and Guangzhou to argue that citizens' value preferences have obvious regional differences, and it is difficult to reach an agreement between government-constructed values and citizens' value preferences[25]. Jiang and Wang (2020) compared and analyzed the similarities and differences of different social media platforms and found that different social media should achieve complementarity and cooperation in responding to public opinion on public emergencies to speed up the rate of information dissemination. Meanwhile, according to the characteristics of different social media platforms, the platform differences were used to improve the efficiency of information dissemination[26]. Sun (2021) conducted a big data analysis of government Weibo in 228 cities and concluded that resource factors play a major role in the development of urban government Weibo[21].

3. Data Sources and Research Methods

3.1 Data Sources

Since the launch of Hunan Taoyuan County's official Weibo "Taoyuan.com" in 2009, major Chinese provinces and cities have set up "Government Weibo". In Sina Weibo, the most widely used social media in China, official Weibo headed by provincial governments are set up to regularly publish some social hotspots and interact with netizens, forming a new channel for government administration and propaganda under social media, which is gradually being used by more citizens. Using big data to analyze the development of China's government Weibo requires the support of a large amount of data. To fully reflect the development of China's government Weibo, this study integrates web crawler technology to collect Weibo data from 31 provincial official websites in China (including Shandong, Hebei, Shaanxi, Neimenggu, Liaoning, Jilin, Heilongjiang, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Hubei, Henan, Hunan, Guangdong, Guangxi, Hainan, Sichuan, Guizhou, Yunnan, Xizang Shanxi, Gansu, Qinghai, Ningxia, Xinjiang, Beijing, Tianjin, Shanghai and Chongqing 31 provinces in total), covering official government Weibo data from January 2018 to April 2022, including multiple dimensions such as Weibo content, public engagement (number of likes, retweets, etc.), and user attributes. Sina Weibo has reserved the data crawling interface to ensure the legitimacy of the data source.

Specifically, this study uses the PyCharm compiler tool for data collection work. The Scrapy framework is used in the python 3.9 environments to implement the collection of website data and extraction of structural data. The collected data contains Weibo posting time, posting content, public likes, retweets, comments, user's gender, Weibo name, etc. By May

2022, a total of 366,124 original postings were collected from 31 provincial government Weibo across China, and the content contained more than 50 million characters. The Weibo content was preprocessed according to standard practices[13, 27]. In this study, symbols, numbers, punctuation marks, stop words, symbols shorter than 3 characters, punctuation marks, and punctuation marks with less than 10 or more than 100,000 occurrences in the corpus were included in the stop word database, which was done here using the jieba stop word module in python.

3.2 Methodology

The data analysis in this study is mainly on unstructured textual data. The main purpose of this study is to analyze the textual data of 31 provincial government Weibo to understand the current situation and hot issues of the development of the Chinese government Weibo. The text analysis is done using python software, running in python 3.9. First, the data processing in this study is based on the pandas module of python, which is a NumPy-based tool that provides effective data recognition and processing functions. Second, this study is based on a natural language processing model to classify the sentiment value of the text, as well as lexical classification. The sentiment value calculation of the data is mainly implemented through the python open-source module snownlp, which generates the original data set for subsequent analysis. Again, the processed data are used to generate sentiment value analysis, word cloud word frequency analysis, and semantic network graphs. Finally, the LDA topic clustering and visual analysis of the data are realized by using the LDA model, and each topic represents one kind of topic.

4. Research Findings and Discussion

4.1 The growing trend of Government Weibo

In the past few years, government Weibo has experienced an emergence and rapid growth and then gradually stabilized. **Fig. 1** shows the trend of the number of posts from January 2018 to May 2022. The figure shows that the number of posts from government Weibo soars after 2020, and the posting volume rises from an average of about 2,000 posts per month in 2019 to an average of about 7,000 posts per month in 2020, a threefold increase. Since then, the number of posts on China's government affairs Weibo has also continued to rise, with an average of more than 13,000 posts in the first quarter of 2022. Data shows that after the outbreak of COVID-19 in early 2020, the Chinese government paid more attention to the spreading role of social media, and made more use of social media to release epidemic information to guide people to prevent the spread of the epidemic. The number of government announcements using social media has increased significantly, which reflects the shift in the Chinese government's governance model in the context of COVID-19. In addition, there is a cyclical nature to government Weibo posting, with a decline in the number of posts in months with more holidays, such as February each year.

High interactivity is the most distinctive feature of social media, so it is also possible to glimpse another aspect of its influence from the perspective of user interaction and feedback. For Weibo, the number of likes is an important indicator to assess user activity, and **Table 1** shows the total number of posts by province and the number of likes by citizens, with huge differences in the intensity of social media use by province. In general, developed provinces and cities (e.g., Beijing, Shanghai, Shandong, etc.) have a high number of posts and interactions, with Shanghai averaging over 80 likes per tweet. Some provinces (e.g. Hainan,

Ningxia, Guizhou, Guangxi, etc.) have a low number of posts and likes. In general, after the number of posts on government affairs microblogs in each province exceeds 10,000, the number of interactions will increase significantly. However, some provinces are exceptions, Liaoning province has a high number of posts, but the number of interactions is very small, with an average of only 1.78 interactions per Weibo.

Similarly, we found that provinces with richer tourism resources (e.g., Chongqing, Sichuan, Xinjiang, etc.) have high postings and attention to governmental Weibo. This suggests that changes in government management concepts can be influenced by the level of local economic development. Different provincial governments have different levels of recognition of social media, resulting in regional differences. Likewise, government Weibo is increasingly valued by governments as a window for external presentation. At the same time, modern technology can help governments use various electronic media to disseminate information to citizens so that citizens can understand government developments and can obtain various information in time to make accurate decisions[28].

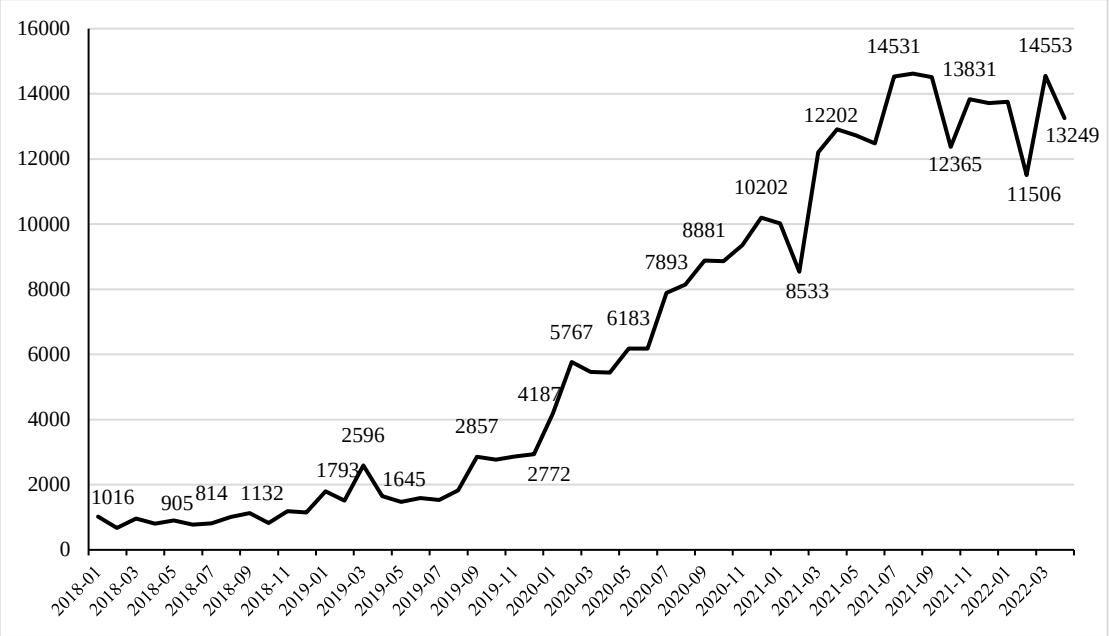

Fig. 1. Time distribution of the number of posts on government Weibo

Table 1. Number of posts and Attitudes by province

Province	posts	Attitudes	Province	posts	Attitudes	Province	posts	Attitudes
Anhui	14,321	385,438	Jiangxi	14,464	73,505	Tianjin	13760	58,982
Beijing	13,614	1,035,424	Henan	1,100	19,616	Jiangsu	14657	146,808
Fujian	6,877	103,960	Liaoning	13,102	23,379	Xizang	9531	18,275
Gansu	14,108	220,509	Ningxia	2,763	624	Xinjiang	14455	273,052
Guangdong	8,532	99,499	Guizhou	4,778	2,013	Yunnan	15044	222,650
Hainan	939	17,599	Qinghai	8,466	21,003	Zhejiang	14310	458,527
Hebei	5,979	16,552	Shandong	14,608	129,456	Guangxi	5365	1,667
Heilongjiang	9,435	361,806	Shanxi	15,498	62,434	Chongqing	13618	833,200
Hubei	13,869	510,343	Shaanxi	12,518	137,013	Jilin	15690	1,159,561
Hunan	9,013	8,757	Shanghai	14,174	1,208,750	Sichuan	15125	272,909
Neimenggu	14,448	31,719						

4.2 Word cloud and word frequency analysis

The steps of analyzing the text data of government Weibo in this section are as follows: First, the data of more than 360,000 government Weibo postings are imported into the python language program. Second, clean the data. For example, some of the postings are just emoji without text, so they will be removed. Third, construct the word separation engine. The word splitting engine in this paper uses Jieba word splitting and manually constructs a comprehensive dictionary containing government actions, government documents, etc., which improves the accuracy of word splitting. Fourth, a corpus is constructed and correlation analysis is performed.

Table 2 shows the top 30 high-frequency words in the text analysis. These high-frequency words reflect the main content of government Weibo to a certain extent. Among the top ten high-frequency words, 5 are directly related to COVID-19, and 3 are indirectly related. This indicates that the Chinese government Weibo has increased its coverage of COVID-19 in recent years, becoming a new channel for the prevention and control of the epidemic in China. After excluding the epidemic-related terms, government Weibo focus on economic development, people's livelihood, and other aspects, such as urban construction, enterprise development, etc., and also includes livelihood-related employment, social security, etc.

Table 2. Word frequency analysis

Ranking	word	Freq.	Ranking	word	Freq.	Ranking	word	Freq.
1	门诊 (Outpatient)	270,673	2	医院 (Hospital)	196,561	3	病例(Cases)	138,270
4	发热(Fever)	108,464	5	疫情 (Epidemic)	106,063	6	医生(Doctors)	105,581
7	工作(Jobs)	104,141	8	开放(Open)	100,090	9	采样 (Sampling)	100,081
10	确诊 (Diagnosis)	95,308	11	数量 (Quantity)	94,877	12	新增 (Additional)	70,193
13	发展 (Development)	70,184	14	发布(Posted)	67,505	15	防控 (Prevention and Control)	65,799
16	境外输入 (Offshore input)	54,802	17	人民(People)	54,297	18	地区(Region)	50,839
19	肺炎 (Pneumonia)	49,936	20	新闻(News)	49,737	21	建设 (Construction)	48,844
22	服务(Services)	47,131	23	企业 (Company)	45,243	24	无症状(No symptoms)	43,152
25	核酸检测 (Nucleic acid testing)	39,132	26	健康(Health)	37,175	27	隔离(Isolation)	34,036
28	保障 (Protection)	33,316	29	医学观察 (Medical Observation)	30,536	30	文化(Culture)	27077

Fig. 2 word cloud diagram shows more keywords, industry, innovation, culture, agriculture, etc. are also focused on by government Weibo, as well as some information disclosure related to the government's operation, such as meetings, international exchange activities, etc. The content of these new governmental media is of general interest to users, and these contents are resonant, such as city identity, people's livelihood, and welfare, and major projects are likely to get the most attention.

We use the constructed corpus to generate co-occurrence matrix nodes and use gephi to produce a semantic network graph. **Fig. 3** shows that in the semantic network, the terms related to the epidemic link most of the keywords and become is the core of the propaganda of the government Weibo, which is marked with orange nodes in the graph. The green nodes are also the focus of government Weibo, and this part tends to be more in the field of people's livelihood, including medical care, environment, and services. The keywords in the purple nodes are related to the field of economic development, focusing more on innovation, production, energy, education, and so on. Finally, the red nodes represent other directions that government Weibo focuses on, such as temperature, culture, tourism, etc. In **Fig. 3**, different nodes are related to each other to form a complete collinear matrix, and the degree of mutual influence between different nodes is determined by the width of the connection. We found that key nodes related to the epidemic will have a huge impact on other fields. In addition, work and development are the most critical nodes other than the epidemic. These keywords illustrate that government Weibo is similar to squares in urban centers, and their influence lies in their ability to deliver information and spread facts in a short period time, especially in the face of sudden public events, which can quickly gather the attention of the masses, create hot spots for public opinion and guide the direction of public opinion.

Fig. 2. Word Cloud Chart

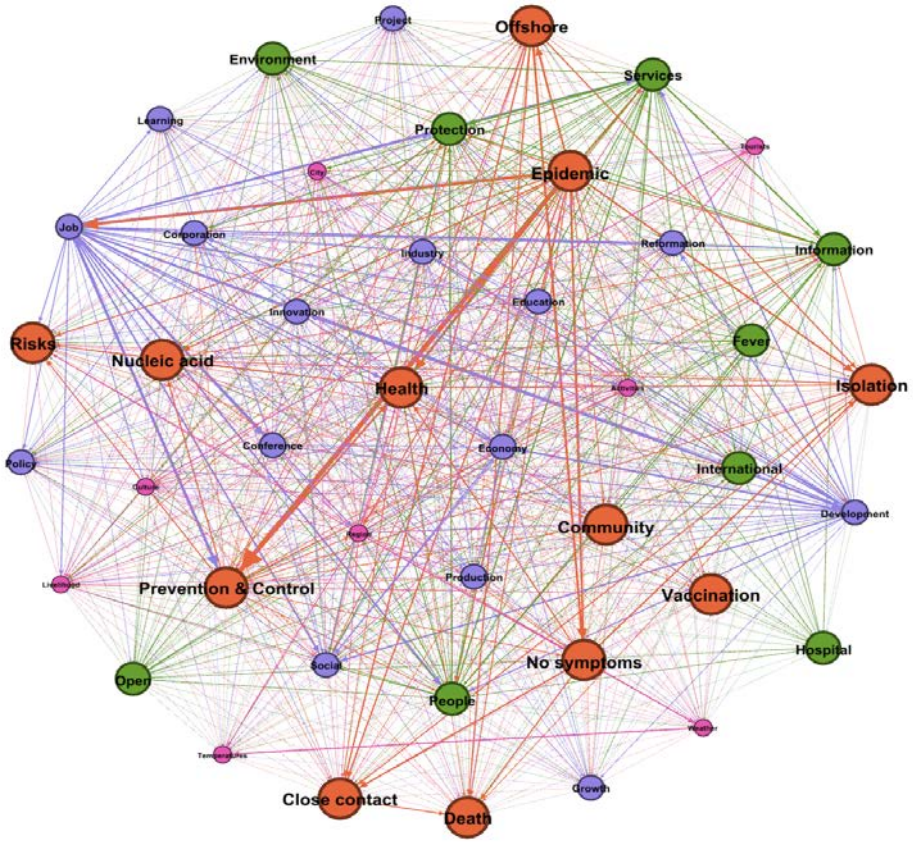

Fig. 3. Semantic Network Graph

4.3 Content Sentiment Value Analysis

This section of the study analyzed users' written comments using a natural language processing (NLP), which was implemented in this study using the python library SnowNLP. The content sentiment value of the content in each document was estimated by extracting the content of each government tweet and identifying a set of key themes from the written document database. Sentiment value scores greater than 0.7 are defined as positive, scores greater than 0.4 and less than 0.7 are defined as neutral, and scores less than 0.4 are defined as passive.

The aggregated results are presented in data visualization, as shown in **Fig. 4**, more than 60% of the content is positive and 31.02% of the content is passive. With the popularity of social media, sentiment analysis technology is more used to identify the emotional trend of topic initiators, and to judge or mine the value of the topic, so as to analyze relevant public opinion. **Fig. 4** also reflects the increase in reports on the epidemic in government microblogs. Although it will increase the amount of attention of the masses, it will also lead to the accumulation of negative emotions. This kind of negative emotion will make the masses too nervous, and to a certain extent will aggravate the contradiction between the masses and managers, which is not conducive to the realization of the government's epidemic control goals.

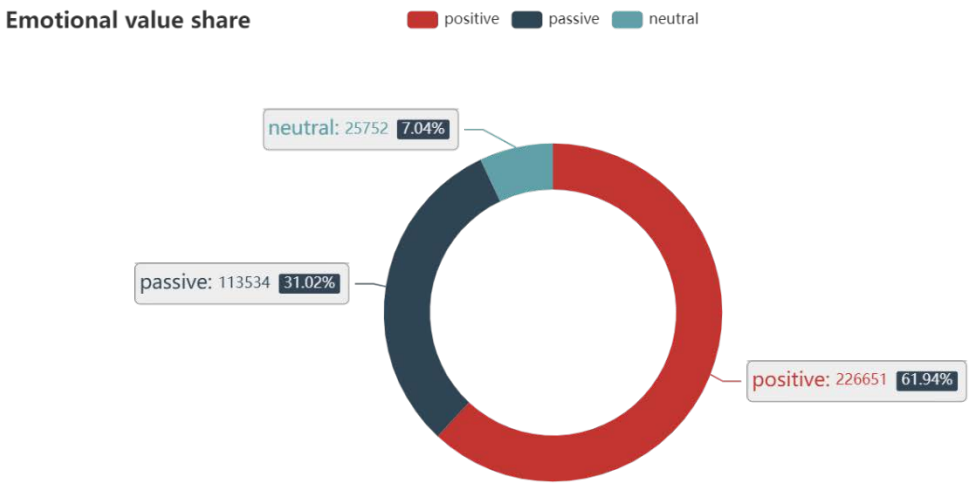

Fig. 4. Emotional value share

4.4 Topic model analysis

We first use the topic model[5] based on the content sentiment values obtained in Section 4.3 and calculate the topic proportion by analyzing the topics that each word in the Weibo content may be related to. Then we use the method of unsupervised learning model of Latent Dirichlet Allocation (LDA) to analyze the content of government Weibo and calculate the relevant topics. LDA, a natural language processing tool for structural analysis of semi-structured text, identifies several themes latent in the policy text and calculates the probability share of each theme to appear in the text. The probability ratios of each theme reflect the relative importance of each orientation in the definition of the theme. LDA provides technical support for researchers to measure topic definition based on a large amount of text data. The model requires the researcher to decide the number of subjects at their discretion, and according to Roberts et al. (2015), the optimal number of subjects should be selected by balancing exclusivity and semantic consistency, with the number of subjects being optimal between 3 and 100[29].

In this study, 10 topic terms are selected for discussion to analyze the effect of government Weibo through machine learning methods and to discuss the application of big data in government management. The reason for choosing 10 topics is that the content topics of government Weibo are relatively focused, mostly in the areas of the epidemic, people's livelihood, and development. Although more topics will contain more complex insights, they also make the analysis more difficult.

Fig. 5 shows the correlation between themes. When conducting the theme model analysis, the themes with a higher percentage tend to be less related to other themes; conversely, the themes with a lower percentage are more related to other themes. Meanwhile, if the distance between two themes is larger, it means less connection, and the closer distance means more connection. 10 numbers in **Fig. 5** represent 10 themes, which are: 1 represents epidemic, 2 represents epidemic prevention and control, 3 represents weather, 4 represents urban development, 5 represents winter Olympics, 6 represents government activities, 7 represents urban tourism, 8 represents special events, 9 represents work entrepreneurship, and 10

represents People's livelihood area. As can be seen from the figure, themes1,2 have accounted for a very large proportion of government Weibo in recent years, which correlates with China's epidemic prevention and control policies and suggests that the government's response to COVID-19 is directly related to effective virus controls. Similarly, government interaction with citizens on social media can form a defense network against virus transmission at the societal level and lead to greater citizen trust in the government[19]. However, this can also lead to too much focus on the topic of the epidemic at the societal level at the expense of other content, thus inhibiting the healthy development of government Weibo.

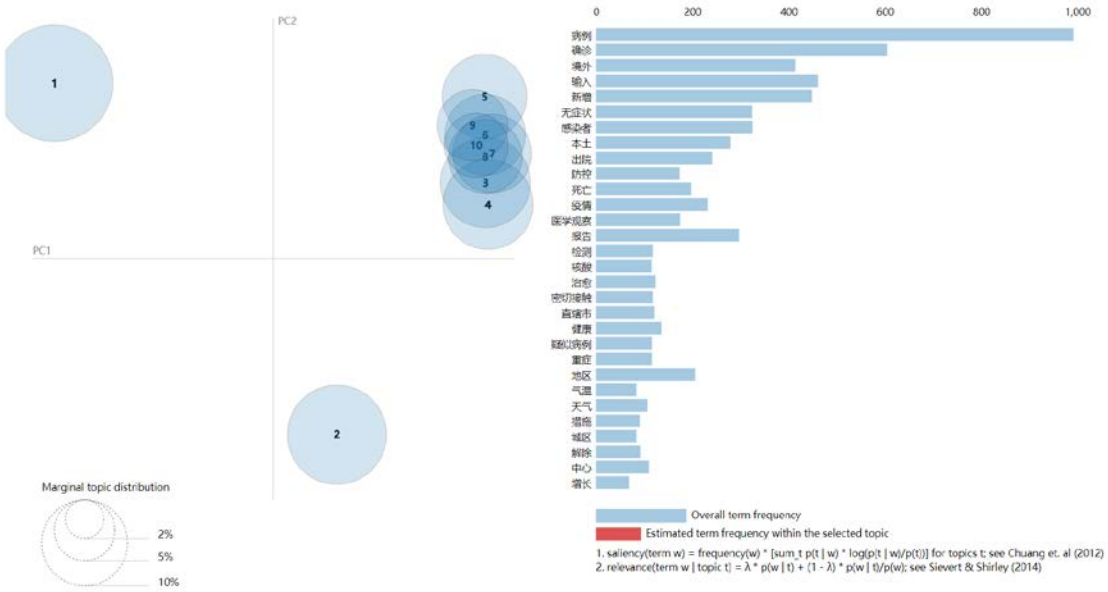

Fig. 5. Intertopic Distance Map

In **Fig. 6-Fig. 7** the top 30 most relevant keywords for topics 1 to topic 3 are listed. The horizontal axis of the figure represents the topic share, and the vertical axis is the 30 high-frequency words for each topic, with longer red bars representing a higher keyword share. In topic 1, for example, among the terms related to the outbreak, Cases, Diagnosis and Offshore input are the three most relevant keywords. In topic 2, prevention and control, health, and detection are the most critical terms. In topic 3, center, weather and temperature are the most critical terms. The graph shows that the LDA model accurately identifies the theme types and aggregates them, which can help us understand the government policy orientation. We found that the keywords of topic 1 - cases and confirmed cases, and the keywords of topic 2 - prevention and control, were all divided into the corresponding topics, which reflected the robustness of the method. Taking topic 1 as an example, under the topic of the epidemic, the content published on government Weibo not only focuses on the domestic epidemic situation in China but also pays attention to overseas epidemic information, which is consistent with the strict control of overseas imports in China's epidemic prevention and control work. This fully reflects the consistency of China's government microblogs with national policies. All in all, the 10 themes are rich in covering many aspects of Chinese government administration. At the same time, there is a significant difference in the percentage of each topic in government Weibo postings. This confirms that the development of China's government Weibo is increasingly being controlled and "guided" by the government. The theme of government Weibo also reflects that the way of government management in the era of big data is no longer

monotonous policy propaganda, but starts from the needs that fit the actual life of the people, so that authoritative and accurate government information can be widely disseminated in the network field and public information dissemination system. The above analysis shows that government Weibo can not only provide a platform for public opinion response but also accept social supervision, which helps to enhance the society's trust in the government. Therefore, proactive disclosure can play the role of the government in actively guiding the direction of public opinion.

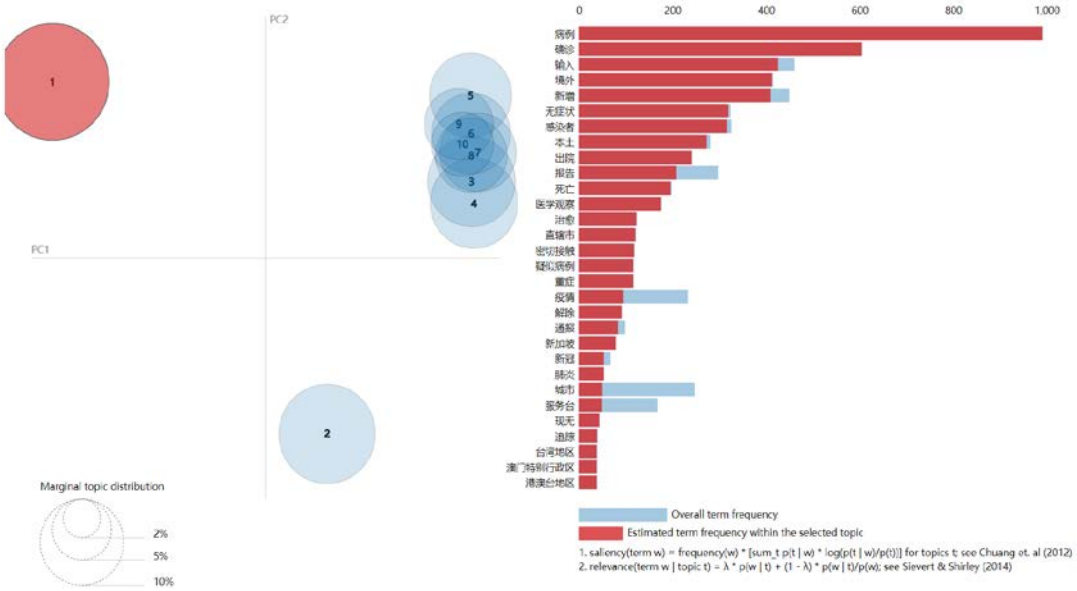

Fig. 6. Top-30 most relevant terms for topic1

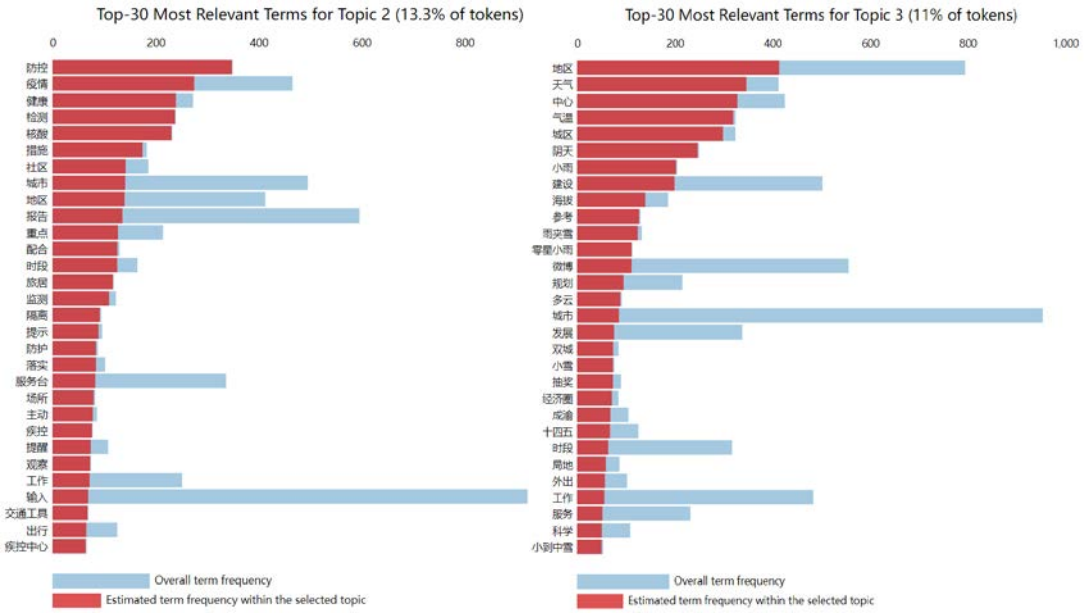

Fig. 7. Top-30 most relevant terms for topic2 and topic3

5. Conclusion

Technological innovations in information communication technology (ICT), especially the rise of social networks such as Facebook, Twitter, and Weibo, have transcended the limitations of traditional digital technologies such as online consultations and forums that separate network users in their use, making it easier for citizens to access government agencies through social media platforms to provide information provided by government agencies through social media platforms. At the same time, governments can also use social platforms such as government Weibo to respond to emergencies in a timely and effective manner and improve social stability. The main work of this paper is to analyze the communication characteristics of government Weibo using text analysis as well as machine learning methods for the big data features of social media to grasp the attribute characteristics and laws of government Weibo. The study finds that, first, government Weibo have entered a booming phase of social media development, with the number of posts as well as their size growing rapidly. Second, in terms of communication content, government Weibo has taken on some of the distinctive features of social media: epidemic, livelihood, and service keywords have become the core, and these are also the main contents that users and communities spread and diffuse, helping the government to build a positive image. Third, from the perspective of natural language learning, the government posting content will be too much involved in the epidemic situations leading to the problem of a high percentage of negative sentiment in government Weibo content. After the LDA model classifies the government Weibo content by topic, epidemic situation and epidemic prevention and control are the two key topics, which overshadow the hotness of other topics, which will inhibit the diversified development of government Weibo.

This study provides some insights into the development of government Weibo in the era of big data: First, government Weibo has both the attributes of government public services and social media. The government Weibo needs to combine the content of their posts with the information needs of the community to attract the attention of users. In the era of social media, the government should pay more attention to the social attributes of government Weibo, communicate more with users, and form a benign feedback mechanism. Government Weibo needs to disseminate more information related to people's livelihoods to ensure the diversity of content. Second, in a social media environment, user experience is very important. The friendly and relaxed narrative style of government Weibo will stimulate users' interest and create a positive new social media environment for government affairs. Moreover, the content published on the provincial government Weibo cannot copy the official news of the superior and should adjust measures according to local conditions. Third, at present, China's government Weibo can be updated in real-time, but different provinces have different levels of awareness of microblogs, and there are huge differences in their development levels. Government Weibo is not a temporary substitute for responding to public opinion on the Internet. It is necessary to ensure the sustainable and healthy development of government Weibo and ensure that can be updated in real-time. The interaction of government Weibo should listen to the voices of netizens, and rely on the public to promote national governance and social affairs management.

This study contributes to the related fields in the following aspects. First, it provides technical support for new developments in technology cognitive theory at the methodological level based on big data text mining. This paper extends the methodological use of related research by using artificial intelligence technologies such as deep learning, natural language interaction, and knowledge graphs for the study of government Weibo. Second, most of the existing studies focus on typical case studies, supplemented by multiple case comparisons, with few empirical analyses based on massive Weibo data. This paper is an exploratory study,

coupled with big data collection and analysis, which can reflect the overall situation and break through the limitations of previous studies on similar topics on a matter-of-fact basis. This paper also has a limitation for research on the formation mechanism, which will supplement provincial data for analysis in the follow-up research.

Acknowledgement

This work was supported by the Pai Chai University research grant in 2022.

References

[1] Tang, G. and Lee, F. L., "Facebook use and political participation: The impact of exposure to shared political information, connections with public political actors, and network structural heterogeneity," *Social science computer review*, 31(6), 763-773, 2013. Article (CrossRef Link)

[2] Tang, Z., Miller, A. S., Zhou, Z. and Warkentin, M., "Does government social media promote users' information security behavior towards COVID-19 scams? Cultivation effects and protective motivations," *Government Information Quarterly*, 38(2), 101572, 2021. Article (CrossRef Link)

[3] Criado, J. I. and Villodre, J., "Revisiting social media institutionalization in government. An empirical analysis of barriers," *Government Information Quarterly*, 39(2), 101643, 2022. Article (CrossRef Link)

[4] Criado, J. I., Sandoval-Almazan, R. and Gil-Garcia, J. R., "Government innovation through social media. Elsevier," *Government Information Quarterly*, 30(4), 319-326, 2013. Article (CrossRef Link)

[5] Blei, D. M., Ng, A. Y. and Jordan, M. I., "Latent dirichlet allocation," *Journal of machine Learning research*, 3, 993-1022, 2003. Article (CrossRef Link)

[6] Huang, J. and Liu, J., "The Development of Digital Governance in Europe and America and Its Enlightenment to China," *Chinese Public Administration*, 06, 36-41, 2019. Article (CrossRef Link)

[7] Janssen, M. and van den Hoven, J., "Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?," *Government Information Quarterly*, 32(4), 363-368, 2015. Article (CrossRef Link)

[8] Mergel, I., Rethemeyer, R. K. and Isett, K., "Big data in public affairs," *Public Administration Review*, 76(6), 928-937, 2016. Article (CrossRef Link)

[9] Guo, X., Li, B. and Li, X., "A Study on Synergistic Development of Innovative Public Management and Economic Growth Based on Big Data," *Mobile Information Systems*, 2022. Article (CrossRef Link)

[10] Rogge, N., Agasisti, T. and De Witte, K., "Big data and the measurement of public organizations' performance and efficiency: The state-of-the-art," *Public Policy and Administration*, 32(4), 263-281, 2017. Article (CrossRef Link)

[11] Kowalski, R., Esteve, M. and Jankin Mikhaylov, S., "Improving public services by mining citizen feedback: An application of natural language processing," *Public Administration*, 98(4), 1011-1026, 2020. Article (CrossRef Link)

[12] Solka, J. L., "Text data mining: theory and methods," *Statistics Surveys*, 2, 94-112, 2008. Article (CrossRef Link)

[13] Anastasopoulos, L. J. and Whitford, A. B., "Machine learning for public administration research, with application to organizational reputation," *Journal of Public Administration Research and Theory*, 29(3), 491-510, 2019. Article (CrossRef Link)

[14] Hogenboom, F., Frasincar, F., Kaymak, U., De Jong, F. and Caron, E., "A survey of event extraction methods from text for decision support systems," *Decision Support Systems*, 85, 12-22, 2016. Article (CrossRef Link)

[15] Bertot, J. C. and Jaeger, P. T., "The E-Government paradox: Better customer service doesn't necessarily cost less," *Government Information Quarterly*, 2(25), 149-154, 2008. Article (CrossRef Link)
[16] Mergel, I., "A framework for interpreting social media interactions in the public sector," *Government information quarterly*, 30(4), 327-334, 2013. Article (CrossRef Link)
[17] De Vries, H., Bekkers, V. and Tummers, L., "Innovation in the public sector: A systematic review and future research agenda," *Public administration*, 94(1), 146-166, 2016. Article (CrossRef Link)
[18] Bonsón, E., Torres, L., Royo, S. and Flores, F., "Local e-government 2.0: Social media and corporate transparency in municipalities," *Government information quarterly*, 29(2), 123-132, 2012. Article (CrossRef Link)
[19] Mansoor, M., "Citizens' trust in government as a function of good governance and government agency's provision of quality information on social media during COVID-19," *Government Information Quarterly*, 38(4), 101597, 2021. Article (CrossRef Link)
[20] Stieglitz, S. and Dang-Xuan, L., "Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior," *Journal of management information systems*, 29(4), 217-248, 2013. Article (CrossRef Link)
[21] Sun, Z. and Zheng, Y., "City Government Micro-Blogging Development and Driving Factors in China—Based on the Combination of Big Data and Small Data Analysis for 228 Cities(2011—2017)," *Journal of Public Management*, 18(01), 77-89+171, 2021. Article (CrossRef Link)
[22] Xiao, M. and Guo, S., "The Role of Online Public Opinion and Government Microblogs in the Modernization of National Governance," *Administrative Tribune*, 21(04), 5-10, 2014. Article (CrossRef Link)
[23] B. Lu, S. Zhang, W. Fan, "Social representations of social media use in government: An analysis of Chinese government microblogging from citizens' perspective," *Social Science Computer Review*, 34(04), 416-436, 2016. Article (CrossRef Link)
[24] T. Meng, S. Zheng, "Information, Communication and Influence: Government Social Media in Internet Governance——An Exploratory Study through Combining Big Data and Small Data Analysis," *Journal of Public Administration*, 29-52+205-206, 2017. Article (CrossRef Link)
[25] Wang, X. and Wang, Z., "Government-Civilian Interaction, Public Values and Government Performance Improvement——An Empirical Study of Government Micro Blogs of Beijing, Shanghai and Guangdong," *Journal of Public Management*, 14(03), 31-43+155, 2017. Article (CrossRef Link)
[26] J. Jiang, W. Wang, "Research on Government Douyin for Public Opinion of Public Emergencies: Comparison with Government Microblog," *Journal of Intelligence*, 100-106+114, 2020. Article (CrossRef Link)
[27] Grimmer, J. and Stewart, B. M., "Text as data: The promise and pitfalls of automatic content analysis methods for political texts," *Political analysis*, 21(3), 267-297, 2013. Article (CrossRef Link)
[28] Purwanto, A., Zuiderwijk, A. and Janssen, M., "Citizens' trust in open government data: a quantitative study about the effects of data quality, system quality and service quality," in *Proc. of The 21st Annual International Conference on Digital Government Research*, pp. 310-318, 2020. Article (CrossRef Link)
[29] Chuang, J., Roberts, M. E., Stewart, B. M., Weiss, R., Tingley, D., Grimmer, J. and Heer, J., "TopicCheck: Interactive alignment for assessing topic model stability," in *Proc. of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 175-184, 2015. Article (CrossRef Link)

Xiang Li received the M.A. degree in Economics from Ningxia University, China, in 2012 and received the Ph.D. in Economics from Pai Chai University in 2022, Republic of Korea. He has been working at School of Business, Heze University since 2012. His research interests include economic development, policy assessment, income inequality and big data analysis.

Xiaoqin Guo received the M.A. degree in Public Administration from Shanghai Normal University, China, in 2013, and received the Ph.D. in Public Administration from Pai Chai University in South Korea in 2022. She has been working at College of Politics and Law, Heze University since 2013. Her research interests include collaborative governance, big data analysis, digital government research.

Soo Kyun Kim received Ph.D. in Computer Science & Engineering Department of Korea University, Seoul, Korea, in 2006. He joined Telecommunication R&D center at Samsung Electronics Co., Ltd., from 2006 and 2008. He is now a professor at Department of Computer Engineering at Jeju National University, Korea. Dr. Kim has published many research papers in international journals and conferences. Dr. Kim has been served as Chairs, program committee or organizing committee chair for many international conferences and workshops; Chair of ICCCT'11, ITCS'10, HumanCom'10, EMC'10, ICA3PP'10, FutureTech'10, ACSA'09, Em-Com'09, CSA'09, CGMS'09, ISA'09, SIP'08, FGCN'08 and so on. Also Dr. Kim is guest editor of the International Journal of "IET Image Processing" and "Multimedia Tools and Applications". His research interests include multimedia, pattern recognition, image processing, mobile graphics, geometric modeling, and interactive computer graphics. He is a member of ACM, IEEE, IEEE CS, KACE, KMMS, KKITS and KIIT.

Hyukku Lee received Ph.D. in School of Economics of Peking University, Beijing, China, in 2008. He has worked at North East Asia Research as a Principal Researcher in the Research Center of China. And he also joined Korea Institute of Economic Policy(KIEP) as an Associate Research Fellow in Team of China. Now He is a professor at Department of East-Asia Studies, Graduate School of Pai Chai University since 2011. His research interests include Institutional Economics and International Trade and Economics, etc.