



Government Affairs Microblog Investigation and Research

Lie Li^{1,a}, Wenxi Li^{1,b}, Weijing Xu^{1,*}

¹ Wuhan University of Technology

^a2513618311@qq.com, ^b2479156370@qq.com, *1052798853@qq.com

Abstract. In recent years, more and more users have joined online social media groups, and the openness and transparency of social platforms have led to the wide dissemination of information, including government affairs microblogs. Government affairs microblogs have become an important way for people to understand government administration information and government effects. Under the government affairs microblogs, the people's meager comments on it are a form of expression of the public's recognition of the government, and it is an electronic word-of-mouth that has gradually become popular. This study mainly explores the impact of whether the factor of public comments and blog posts are consistent with the factor of blog posts on people's emotions under government microblogs, first creatively establishes the concept of consistency between democratic comment content and blog post content, and then uses LDA theme extraction, text analysis and sentiment analysis to collect the data of 20 government microblogs in 5 cities for verification. The final result show that the influence of the consistency between the factor of public comments and the factor of blog posts under the government's microblogs is positive. The result of this study have an optimal effect on the operation of government microblogs, allowing the government to interact more cordially with the people, and allowing the people to better understand the government's policies and more recognizing the government's governance effects.

Keywords: government affairs microblog; Online reviews; LDA modeling; Sentiment analysis

1 Introduction

With the continuous development of the Internet, the interaction between the people and the government on the online platform is becoming more and more important, and the new e-government platform represented by the government microblogs is one of the key platforms. With these emerging platforms, the activities between the government and the people can be spread on a larger scale. The emotion of people's comments under the government affairs microblogs are an indicator to measure the public's recognition of the attitude of the city and government work. In 2018, the Opinions of the

© The Author(s) 2023

S. Yacob et al. (eds.), *Proceedings of the 2023 7th International Seminar on Education, Management and Social Sciences (ISEMSS 2023)*, Advances in Social Science, Education and Humanities Research 779,

https://doi.org/10.2991/978-2-38476-126-5_30

General Office of the State Council on Promoting the Healthy and Orderly Development of New Media for Government Affairs pointed out: "New media for government affairs is an important channel for the Party and the government to connect, serve and unite the masses in the era of mobile Internet, an important means to accelerate the transformation of government functions and build a service-oriented government, an important way to explore new models of social governance and improve social governance capabilities."

An important research direction in text analysis is sentiment analysis. The process of generating network emotions is a very complex process, which is closely related to the law of network information dissemination and has a far-reaching and profound social background. In terms of sentiment analysis, many achievements have been accumulated domestic and foreign, including emotional classification and emotional polarity analysis. There are two main techniques in emotional polarity analysis, one is the emotional polarity analysis using the emotional dictionary as the reference library, and the other is the emotional polarity analysis that uses the training document set in a specific region as the source of text polarity knowledge, and then establishes the emotional polarity analysis recognized by the text classification model. In the past research, the emotional polarity analysis model established by Ma Chao and Yu Hui has achieved remarkable results¹. Sun Lingling established a text classification recognition model based on LSTM network in order to extract the emotions of people's comment texts when exploring the satisfaction and attention of popular topics, which can completely obtain the unstructured context information of comment texts while avoiding too much reliance on the construction of emotional dictionaries.² This study uses the second method, using Baidu PaddlePaddle as the framework, to analyze the emotional polarity of comments under government affairs microblogs.

In the early days of China, there has been emotional analysis of microblogs comments, but their background is mostly based on ordinary microblogs or sudden public events. An Lu et al. have conducted sentiment analysis on microblogs comments under public health emergencies to study the co-evolution of public opinion and public emotions. In view of the "Tianjin Explosion" case, Zhao Xiaohang replaced the people's emotions with microblogs comment emotions, and conducted a comprehensive analysis of the information release effect and real-time feedback of public opinion. It can be seen that although there are many studies in the Scholarly Community, there is no influencing factor on the emotions of government affairs microblogs comments in daily situations. So we focus on the impact of the consistency of government affairs microblogs articles and comments on people's emotions.

Based on the past literatures, this study mainly establishes a conceptual model of the consistency between the factor of public comments and the factor of blog posts, explores the impact of whether the factor of public comments and blog posts is consistent under government microblogs on people's emotions, and collects the data of 20 government microblogs in 5 cities for corresponding verification.

2 Research Assumptions and Models

2.1 Consistency between blog topics and comment topics

In the past, some studies had objectively extracted the themes of government microblogs through the LDA topic recognition model, and then summarized the themes into blog topics based on the city image theory.

Scholars have confirmed that different topics of blog posts have an impact on the emotional nature and polarity of public comments. By analogy, this study extracts the topics for comments while extracting the topics for government microblogs blog posts, exploring the impact of the consistency of blog post issues and comment topics on people's emotions.

Li Meihua has proposed a CNN-Pay model for sentiment classification in blog posts and comment data, predicting the curve of sentiment value over time, and then analyzing people's emotional trends³. Zhang Xueting and Liu Yindi analyzed the emotional structure and influencing factors of news audiences from a social perspective based on the text of microblogs comments, reflecting the negative emotional mapping of the public's trust in social media⁴. Building on previous researchs, this paper proposes hypotheses:

The consistency of government microblogs topics and comment topics has an aggressive impact on public comment emotions.

3 Study data and variables

3.1 Data collection and preprocessing

In this study, an octopus collector was used to crawl representative government microblogs from January 31, 2021 to January 31, 2022 in Beijing, Shanghai, Guangzhou, Shenzhen and Wuhan, and obtained microblogs blog posts and corresponding comment topics. Table 1 shows the list of microblogs for each city.

Table 1. Government affairs microblogs in various cities

City	Microblogs
Beijing	Beijing Release, Safe Beijing, Weather Beijing, Culture and Travel Beijing
Shanghai	Police Direct-Shanghai, Shanghai Metro, Shanghai Public Security, Shanghai Weather
Guangzhou	China Guangzhou Public Security, Guangzhou Weather, Guangzhou Public Security, Guangzhou Culture, Radio, Television and Tourism Bureau
Shenzhen	Shenzhen Public Security, Shenzhen Weather, Shenzhen Microblogging Office, Shenzhen Health Commission
Wuhan	Wuhan Public Information, Wuhan Fire, Safe Wuhan, Wuhan Tourism

After sorting, a total of 634036 comment data was crawled, invalid and duplicate comments were deleted, and all comment data of the unified main post were merged to obtain a total of 135711. Use the exact pattern of the jieba library in Python to perform word segmentation operations on corpus. Sample text after jieba participle is shown in Table 2.

Table 2. Weibo comment text after jieba word segmentation

Example	Word segmentation results
Original sentence	[#6-year-old girl trapped outside the 4th floor window hanging in the anti-theft window#Police fire rapid rescue] On January 14, a 6-year-old girl in Wuhan was at home alone, climbing on the windowsill to look for her mother, but accidentally got stuck in the anti-theft window, and most of her body was suspended in the air. After the police and firefighters arrived and broke in, police officer Tu Long climbed on the windowsill to support the girl, and the firefighters cut open the security window and carried the child out. Police remind: Parents try not to leave their children alone at home to avoid accidents.
Cut the word result	['Girl', 'Snap', 'Window', 'Hanging', 'Outside', 'Police', 'Fire', 'Fire', 'Rescue', 'One', 'Home', 'Girl', 'Window Sill', 'Outward', 'Look', 'Mom', 'Careless', 'Stuck In', 'Security Windows', 'Most of the Half', 'Body', 'Hanging in the Air', 'Dangerous', 'Police', 'Firefighter', 'Fire Speed', 'Rush', 'Breaking Door', 'Police', 'Tu Long', 'Window Sill', 'Holding'], 'Girl', 'Firefighter', 'Cut', 'Security Windows', 'Be Careful', 'Girl', 'Say', 'Finally', 'Police', 'Reminder', 'Parents', 'Children', 'Alone', 'Stay Home', 'Accident', 'Gao Leizhang', 'Ziwei', 'Serve the People', 'Story']

After the word segmentation, the letters, numbers and punctuation words in the corpus are first removed by regular expression, and then the corpus is cleansed by using the HIT stop word list. Then, the cleaned corpus was analyzed, the words with the top 100 frequencies were obtained, and the words with a large number of times but irrelevant to the topic and interfered with the later research were screened out, such as words such as "unsealing" and "forwarding", and they were added to the stop words for secondary data cleaning, and finally the corpus for analysis was obtained.

3.2 Summarize blog post topics and comment topics

ZhangZhaoyu et al. ⁵used the LDA theme extraction model to explore the relationship between article topic and user behavior. This study followed this method by using the LDA model to extract blog post topics and comment topics, and then categorized them into blog post topics and comment topics. When using LDA to extract topics, this study uses the method of simulating manual evaluation results to determine the optimal number of topics to ensure the quality of theme templates. Simulated manual assessment can be divided into direct assessment and indirect assessment, direct assessment of the

consistency⁶ of the calculation theme, and indirect assessment of calculation confusion⁷. For this study, the lower the level of confusion, the clearer the extracted topics. However, after reaching a certain amount, the larger the number of topics, the lower the confusion, so you can't just consider the confusion. The higher the topic consistency, the stronger the differentiation between different subjects, because the consistency can also reflect the interpretability⁷ of the model, so at present, consistency is a better topic model evaluation standard than puzzle. In this study, the criterion for evaluating the theme model was consistency, supplemented by confusion.

In the course of the study, the number of topics ranged from 1 to 500, with a span of 25 to calculate the consistency and confusion, and summarize their change patterns. After collation, it was found that the interval with better theme consistency results was (19,24) and the difference with better perplexity results was (10,30). Therefore, in order to make the number of topics with higher consistency and less confusion, the research selects a new range to output the results one by one, and finally determines the number of topics to be 20. Finally, the Ldamodel class modeling in the gensim library was used to obtain the top 10 keywords with 20 topics and weights, as shown in Table 3 and Table 4.

Table 3. Main post subject

topic	keyword
topic0	'Good morning' 'Good morning' 'Cloudy' 'morning and evening' 'dry' 'Sunshine' 'Cloudy' 'Light' 'Rain' 'Working' 'Sunny'
topic1	'New Year' 'Mobile' 'Action' 'recommended' 'Try it' 'Cool' 'Smile' 'Prevention' 'sequentially' 'Follow-up'
...	...
topic18	'Warning' 'Omicron' 'Hour' 'Yellow' 'Signal' 'Forest' 'Dressing' 'Spoiler' 'Happening' 'Connectivity'
topic19	'Cold air' 'Affect' 'Off work' 'Temperature' 'Wind gusts' 'Cooling' 'Cloudy' 'Falling' 'Cold' 'Scam'

Table 4. Comment on the topic

topic	keyword
topic0	'Cold Air' 'CCTV' 'Dry' 'Good' 'Sunshine' 'Morning and Evening' 'Voting' 'Sun' 'Remember'
topic1	'Passenger flow' 'Peak' 'Suggestion' 'Dynamic' 'Yesterday' 'Time slot' 'Live' 'Opinion' 'Feedback' 'Crowding'
...	...
topic18	'Guard' 'Ventilation' 'Delicious' 'People' 'Site' 'Junction' 'Happy New Year'
topic19	'Wanjia' 'Understanding' 'Limited Company' 'People's Police' 'Police People' 'Masses' 'Panda' 'Food' 'Discovery' 'Year of the Tiger' 'New Year' 'Mobile Phone' 'Happy'

After summarizing the theme keywords, we categorized the topics of the blog into seven main topics: life services, transportation and commuting, event festivals, medical

resources, introduction to government affairs, culture, sports and entertainment, and humanistic feelings. Table 5 summarizes the topics.

Table 4. Seven types of microblogs topics.

Attributes	Topic Number	Comments
Life Services	0,16,18,19	0,3,9,15
Transportation Commuting	6,12,13,14,15	1,6,13,14
Event Festivals	1,11,17	8,18,19
Medical Resource	4,8,10	2,11,16
Introduction to government affairs	5,7	10,17
Culture,Sports and Entertainment	3,9	4,7
Humanistic feeling	2	5,12

3.3 Sentiment analysis

In this study, Baidu's PaddlePaddlePaddle was used for sentiment analysis, Baidu's emotional tendency thesaurus included articles and phrases in various fields, which is a high-availability and high-accuracy emotional tendency model, and the emotional tendency of Flypaddle is an efficient sentiment analysis model. The sentiment model was used to analyze the sentiment of the corpus, and the sentiment of the comment set was quantified as a floating-point number between 0 and 1, and the closer the score was to 1, the more positive the emotion.

4 Hypothesis validation

4.1 (1) Variable Definitions

In this study, we intend to test the hypothesis using a linear regression equation, with the independent variable being the consistency of the blog post topic and the review topic, and the dependent variable being the sentiment score. The data description statistics are shown in Table 6 and Table 7:

Table 5. Numeric data description

variable	quantity	mean	standard deviation	minimum	maximum
Attitude	135711	0.874211	0.149594	0.5	0.999747

similar	0.827	0	-	1677.44 9	0.000** *	-	0.1 4	0.14	F=22163.968 P=0.000***
	0.114	0.001	0.375	148.876	0.000** *	1			

Note: ***, **, *Respectively represent the level of 1%, 5%, 10% significance

5 Conclusion and discussion

5.1 Analysis of Results

Based on the above model results, the hypothesis is verified, and the above data analysis concludes:

In the analysis of variance, it has been proved that the consistency of blog post issues and comment topics can have an impact on people's emotions. Combined with the time of data acquisition, when the content of the comment is consistent with the theme of the content posted by the blogger, people show more positive emotions, which indicates that when the official released the content that the public wanted to see at that time period, the people's emotions were more positive. The significance level is 1% and the coefficient is positive, so the assumption holds.

5.2 Policy Recommendations

The content of information released by government affairs microblogs involves the practical interests of a wide range of people, and needs to meet the great information needs of the masses. Especially in today's information age, timely and quantitative information communication between relevant government departments and the general public is quite important, and the establishment of the government affairs microblog platform provides a good information platform for the general public and relevant government departments. This study examines whether the topic categories of blog posts in blog post posting strategies are consistent with the topic categories of public comments, and examines how government microblogs can effectively make people emotionally positive. The results showed that when the blog post topics in the blog post were consistent with the category of people's comment topics, people's emotions were more positive. This shows that when people see the government affairs Weibo release the information that the people want to see recently, they will participate more actively in responding and achieve a good interactive atmosphere

Therefore, it is recommended that local government affairs microblogs should effectively consider recent matters related to people's livelihood when editing the blog content posted by Weibo masters, and publish things that people want to know recently, so as to create a good interactive atmosphere and platform.

References

1. SUN Lingling, HU Yanrong, LIU Hongjiu. Research on user concerns of online brand community based on LSTM-LDA algorithm and IPA analysis[J]. *Journal of Intelligence*, 2021, 40(09): 178-186.
2. Li Meihua, Research on microblogs Public Opinion Analysis Based on Clustering and Deep Learning. (10.27176/d.cnki.gnfy.2022.000814)
3. Xueting Zhang, Yindi Liu, Analysis of the Emotional Structure and Influencing Factors of Inverting News Audiences from the Perspective of Social Trust: Text Mining Based on microblogs Comments. 10.13556/j.cnki.dncb.cn35-1274/j.2021.07.032)
4. ZHANG Zhaoyang, WANG Junling, HUANG Jiani, SHEN Jiayu, WANG Yiyang, ZHAO Rongying. Research on the correlation and evolution of microblog themes and user behavior during major emergencies[J]. *Journal of Information Resources Management*, 2021, 11(02): 28-38. DOI: 10.13365/j.jirm.2021.02.028.)
5. Newman D, Lau J H, Grieser K, et al. Automatic evaluation of topic coherence[C]// *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*. 2010: 100-108.
6. Mimno D, Wallach H, Talley E, et al. Optimizing semantic coherence in topic models[C]// *Proceedings of the 2011 conference on empirical methods in natural language processing*. 2011: 262-272.
7. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures[C]// *Proceedings of the eighth ACM international conference on Web search and data mining*. 2015: 399-408.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

