



# Research on Predicting Social Media User Communication Behavior Based on LIWC from the Perspective of Psychological Language

Chuanlin Huang<sup>1,\*</sup>, Yanxia Lu<sup>2</sup>

<sup>1</sup>School of Information and Business Management, Dalian Neusoft University of Information, Dalian, China

<sup>2</sup>School of Management, Liaoning Normal University, Dalian, China

\*huangchuanlin@neusoft.edu.cn, luyanxia@lnnu.edu.cn

**Abstract.** In major emergencies, social media has accelerated the irrational dissemination of online public opinion. As the most important user information behavior, retweeting indicates the development trend of network public opinion to a large extent. However, existing research has paid little attention to the relationship between psychological language use and communication behavior contained in the information content published on social media. This study theoretically expands the application scenario of psycholinguistics in emergency situations, and also provides reference significance for the government and emergency management departments at all levels to effectively guide public opinion in emergencies. Taking Tianjin Port explosion incident as an example, this paper studies the impact of social media users' psychological process on communication behavior based on text analysis tool LIWC. By constructing VAR vector autoregressive model and conducting Granger causality test, the influential factors of social media communication behavior are determined from the psycholinguistic perspective. Furthermore, impulse response function is used to analyze the propagation behavior dynamically. According to the results of empirical research, emotional process words, perceptual process words, social process words and physiological process words all have a certain predictive effect on the communication behavior of social media users.

**Keywords:** communication behavior, formatting, social media, information content, psycholinguistic.

## 1 Introduction

With the rapid development of social media, as the largest social media information dissemination platform in China, Weibo has gradually become an important way for the public to disseminate and obtain information, and has changed people's thinking and lifestyle to a large extent, becoming an indispensable part of people's life. It has had a significant impact on society, politics, economy, culture and people's daily life. People often share special, shocking or eye-catching news on Weibo, which allows

© The Author(s) 2024

Z. Zhan et al. (eds.), *Proceedings of the 2024 10th International Conference on Humanities and Social Science Research (ICHSSR 2024)*, Advances in Social Science, Education and Humanities Research 858,

[https://doi.org/10.2991/978-2-38476-277-4\\_38](https://doi.org/10.2991/978-2-38476-277-4_38)

people to quickly exchange risk-related information with each other<sup>[1]</sup>. Especially in emergency situations, more and more people use Weibo to send and receive information<sup>[2]</sup>.

There are usually four kinds of users' behavior on social media: tweet, retweet, comment and like. Among them, tweet, retweet and comment can promote the spread of social media information. However, retweet is still one of the most popular functions among users in social networks, and it is also the most important embodiment of social media information transmission behavior. In Weibo, people can choose to retweet messages on their blog space. In this way, the information carried by the message can be quickly spread in the social network. Both domestic Weibo and foreign Twitter are media platforms with large scale, large number of users, and deep influence. More and more scholars are studying the forwarding behavior of domestic and foreign media platforms<sup>[3]</sup>.

Retweet is a very important user behavior that affects online public opinion in social emergencies. The widespread dissemination of information, spread of rumors, and outbreak of public opinion are all closely related to audience information forwarding<sup>[4]</sup>. Therefore, this paper takes retweet as the representative variable of communication behavior to study the predictive factors that affect the public's information communication behavior in social emergencies. In addition, there are many methods to predict users' information communication behavior, but there are relatively few researches on communication behavior prediction based on users' mental process and psycholinguistic characteristics in emergencies. Therefore, this paper takes Weibo as a typical social media platform and Tianjin Port explosion as a typical case. By introducing psycholinguistic variables, VAR regression model is constructed and impulse response graph is drawn to predict the communication behaviors of social media users.

## 2 Related Works

The rapid development of information technology provides two important ways for information dissemination: network information tweet and network information retweet. Among them, network information tweet mainly propagates original information, and network information retweet mainly propagates non-original information<sup>[5]</sup>. When a post catches the attention of the public and is retweeted tens of millions of times, the message spreads quickly on social networking sites. Finding out the key factors that lead to the information retweeting behavior of microblog users can help us understand the internal mechanism of information transmission. More and more scholars begin to pay attention to the information retweeting behavior of microblog users, and use different methods to analyze the behavior of microblog forwarding.

Zhou<sup>[6]</sup> proposed a Weibo audience user influence measurement algorithm IKAGR based on the characteristics of Weibo social platforms and the correlation between participants and reposters in Weibo topics. Ding<sup>[7]</sup> used two different types of unexpected events on Sina Weibo as examples, and used liking, commenting, and forwarding as characteristic indicators to conduct predictive experiments on user interaction behavior.

Feng<sup>[8]</sup> studied the communication rules of government affairs microblog from the perspective of public micro-behaviors, and found that social capital, social trust and behavioral habits would all affect the communication behaviors of Weibo users, thus affecting the communication effect of government affairs microblog. Zhang<sup>[9]</sup> studied the factors influencing the effectiveness of social media user information acquisition under the background of the information epidemic, guided users to engage in effective information acquisition behavior, and further promoted the construction of public crisis information management.

Furthermore, people often share their thoughts, feelings and attitudes towards emergencies by posting or forwarding on social media. By analyzing the psychological process of users in microblog content, users' forwarding behavior can be understood. Liu<sup>[10]</sup> studies the bounded rational psychological behavior of decision makers by using intuitive language set, and believes that language can reflect people's mental world and influence their decision-making behavior to a certain extent. Bai<sup>[11]</sup> takes the data text information left by college students on social media and various online platforms as the research object, and uses natural language processing to conduct mathematical modeling. It introduces a text information processing network TCNN based on convolutional operations.

Word use can tell us far more than just the story, it is a window into the narrator's world. Pennebaker and Stone<sup>[12]</sup> analyzed how people change in their use of 14 text dimensions as a function of age. Kramer and Rodden<sup>[13]</sup> present a large-scale analysis of the content of weblogs dating back to the release of the Blogger program and suggests that blogs vary along five psychologically relevant linguistic dimensions: Melancholy, Socialness, Ranting, Metaphysicality, and Work-Relatedness. LIWC was developed in early 1990s to map psychological and linguistic dimensions of written expression<sup>[14]</sup>, and is now often the preferred automated text analysis method in psychology, and an important choice of natural language processing in computer sciences. From the available LIWC variables, some language dimensions were high relevant for understanding differences in self-presentation styles because they have previously been linked to different coping strategies, such as emotional processes, cognitive processes, social processes.

With the popularity of mobile networks, the number of Weibo users has exploded, and the speed of the spread of emergencies is much faster than ever before. The above literature presents some studies on information diffusion behavior. However, few studies have focused on the use of psycholinguistic language in microblogs. Therefore, using psycholinguistic features to analyze information retweeting behavior has very important practical significance.

### 3 Research Design

According to emotion contagion theory<sup>[15-17]</sup>, when the emotions of others infect individuals, the emotions of individuals will be returned to others, thus strengthening the original emotional state of others. After such repeated reinforcement, some emotions

within the group will gradually achieve inter-group assimilation. Furthermore, the functional and emotional vocabulary that people use can provide important psychological cues to their thought processes, emotional states, intentions and motivations, and has been widely used in many areas such as academic performance, health improvement, and perceptual abilities. Therefore, this paper will discuss the information dissemination behavior of social media users from the perspective of people's psychological process. The research scheme is shown in figure 1.

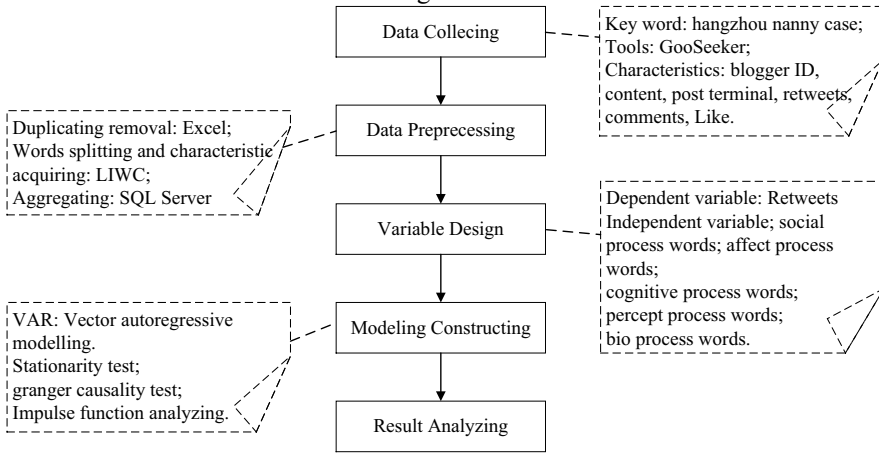


Fig. 1. Research scheme.

The stepwise is as follows:

Step 1: data collecting. Data were collected by the key word with the crawler tool, GooSeeker. The characteristics of microblogs consist of blogger ID, content, post terminal, retweets, comments and likes.

Step 2: data preprocessing. First, duplicating and unrelated posting are deleted. Then, applying LIWC as a tool to analyze the content of microblogs, and we acquired the corresponding characteristic values of each microblog post. Last, aggregating all of the characteristics from SQL Server according to hour.

Step 3: variable design. The topic of this paper is retweeting behavior, so we choose retweet as dependent variable, and psychology process words as independent variable, which consists of social process words, affect process words cognitive process words, percept process words and bio process words.

Step 4: modeling constructing. Retweeting behavior is always affected by microblog of early stage. Therefore, we constructed a VAR model. We first conduct stationarity test in order to confirm the stability of data, and then propose the granger causality test for discussing the cause-effect relationship. Finally we get an impulse function analyzing.

Step 5: result analyzing. According to the result of VAR and impulse function, the influence of main psychological words on the communication behavior of social media users is discussed.

## 4 Empirical Study

On August 12, 2015, a fire and explosion occurred at the dangerous goods warehouse of Ruihai Company in Tianjin Port, Binhai New Area of Tianjin, killing 165 people, missing 8 people, injuring 798 people and damaging 304 buildings, 12,428 commercial vehicles and 7,533 containers. The investigation team of The State Council determined that the fire and explosion accident of the dangerous goods warehouse of Ruihai Company in Tianjin Port "8·12" was a particularly major production safety responsibility accident.

### 4.1 Descriptive Statistics

Using "Tianjin Port explosion" as the keyword, data from August 12, 2015 to October 31, 2015 were retrieved, with a total of 38,841 microblogs. By removing duplicate and irrelevant content, we ended up with 38,812 posts. And then we conduct words segmentation and remove stop words, finally aggregate by hours. We get six variables, and names dependent variable as retweet, independent are social (social process words), affect (affect process words), cognitive (cognitive process words), percept (percept process words) and bio (bio process words).

We figured the sequence diagram of variables, and find that all of variables are related with retweeting behavior. From August 12 to August 13, the day after the incident broke out, it reached the peak of public opinion and dissipated a week later. The sequence analysis is shown in figure 2.

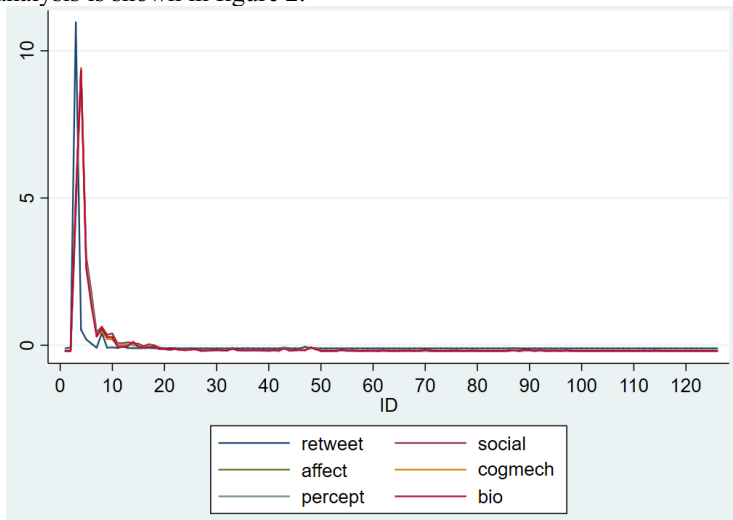


Fig. 2. Sequence diagram.

The descriptive statistics result is shown in table 1.

**Table 1.** Descriptive statistics result.

Stats	Retweet	Social	Affect	Cogmech	Percept	Bio
mean	887.4381	3.707284	3.044794	7.468046	1.292831	1.575326
sd	1916.453	.9374552	.8721357	1.529649	.5292027	.626183
skewness	4.45018	.6073432	.0811503	.1864513	1.371524	1.990856
kurtosis	27.39005	3.063311	3.218645	3.52468	7.887219	10.07006

From the result of table 1, we can see that the difference of variables are very large and is too discrete, in order to avoid inconsistencies in the data unit, we conduct data standardization using the formula:

$$x^* = \frac{x - \mu}{\sigma}$$

Where  $x$  is the original value,  $\mu$  is the mean of the sample, and  $\sigma$  is the standard deviation of the sample data, which is the normalized data. Therefore, we can get a standard normal distribution with a mean of 0 standard deviation of 1.

### 4.2 Correlation Analysis

Correlation analysis is conducted in order to test the correlation between the variables. The correlation coefficients matrix between the variables studied in this paper is shown in table 2. Since each variable is a continuity variable, Pearson correlation analysis is used to obtain the correlation coefficient  $r$ .

$$r = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)$$

**Table 2.** Result of correlation analysis.

	zretweet	zsocial	zaffect	zcogmech	zpercept	zbio
zretweet	1					
zsocial	0.4586*	1				
zaffect	0.5047*	0.9976*	1			
zcogmech	0.4889*	0.9979*	0.9997*	1		
zpercept	0.4819*	0.9990*	0.9996*	0.9997*	1	
zbio	0.5088*	0.9964*	0.9995*	0.9993*	0.9989*	1

From the above table we can see that there is a certain correlation between retweet and other variables, where the significance level is at 0.05.

### 4.3 Regress Analysis

When time series data is not stable, it may cause false correlation or false regression. So we test the stationarity of the data firstly. In this paper, Stata was used for ADF unit root test, all of the values of p are remarkable, which represents all variables are all stationarity.

Using VAR model, the lag time is p, and the formula is as follows:

$$y_t = A_1y_{t-1} + A_2y_{t-2} + \dots + A_p y_{t-p} + \varepsilon_t$$

$y_t$  is a vector of m dimensions.  $A_1, A_2, \dots, A_p$  is the parameter matrix to be estimated.

Here, in order to ensure the balance between lag period and degree of freedom, the lag order p is 5 according to SBIC (Schwarz's Bayesian information Criterion), with minimum information value.

Therefore, the final formula can be expressed as follows:

$$y_t = A_1y_{t-1} + A_2y_{t-2} + A_3y_{t-3} + A_4y_{t-4} + A_5y_{t-5} + A_6y_{t-6} + A_7y_{t-7} + \varepsilon_t$$

We conducted Granger causality test. The model fitting information is as shown in table 3. We can see that all of variables are prominent. Here, SBIC has the minimum value -41.5421 in lag 5.

**Table 3.** Granger causality test (1).

Equation	RMSE	R-sq	F	P > F
zretweet	0.005456	0.9924	331.237	0
zsocial	0.020084	0.9947	474.9778	0
zaffected	0.015507	0.9955	565.6859	0
zcogmech	0.014267	0.9961	644.9017	0
zpercept	0.016249	0.9955	562.3801	0
zbio	0.012951	0.9963	679.9946	0
AIC=-46.358;HQIC=-44.4084;SBIC=-41.5421				

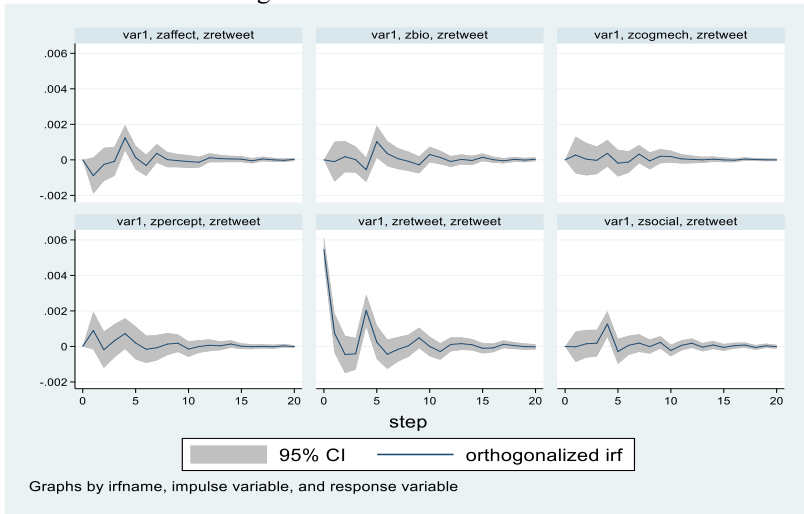
The part result of Granger causality test is as shown in table 4, here, we just take retweet as an example. From the table we can see that retweet is prominent related with itself in lag 1, lag 2 and lag 5. Social process is related with retweet in lag 1 and lag 4. Affect process is related with retweet in lag 1. Percept process is related with retweet in lag 4. Bio process is related with retweet in lag 4. Therefore, social, affect, percept and bio is the cause of retweeting behavior. Furthermore, cognitive process are unrelated with retweeting behavior.

**Table 4.** Granger causality test (2).

zretweet	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
zretweet						
L1.	0.3118692	0.1378354	2.26	0.027	0.0373461	0.5863922
L2.	-0.3291871	0.1173671	-2.8	0.006	-0.5629439	-0.0954302
L3.	-0.0017361	0.0772548	-0.02	0.982	-0.1556023	0.1521301
L4.	-0.0246772	0.0603231	-0.41	0.684	-0.144821	0.0954667
L5.	-0.0932058	0.0348399	-2.68	0.009	-0.1625954	-0.0238162
zsocial						
L1.	0.3048113	0.1364931	2.23	0.028	0.0329616	0.5766609
L2.	-0.1396837	0.1294526	-1.08	0.284	-0.3975108	0.1181434
L3.	0.093534	0.1100481	0.85	0.398	-0.1256459	0.3127138
L4.	-0.2091544	0.0801443	-2.61	0.011	-0.3687757	-0.0495332
L5.	-0.0038038	0.0608562	-0.06	0.95	-0.1250095	0.1174019
zaffect						
L1.	-0.7333394	0.3385061	-2.17	0.033	-1.407533	-0.059146
L2.	0.1975796	0.2577822	0.77	0.446	-0.3158381	0.7109973
L3.	-0.157689	0.2552743	-0.62	0.539	-0.6661119	0.3507339
L4.	-0.0315419	0.2303758	-0.14	0.891	-0.4903751	0.4272913
L5.	0.0197693	0.2141664	0.09	0.927	-0.40678	0.4463186
zcogmech						
L1.	0.0785739	0.3062468	0.26	0.798	-0.5313696	0.6885173
L2.	0.0925279	0.2322881	0.4	0.692	-0.3701139	0.5551697
L3.	-0.1634524	0.197739	-0.83	0.411	-0.5572837	0.2303789
L4.	0.1781663	0.22127	0.81	0.423	-0.2625311	0.6188636
L5.	-0.0783963	0.1870032	-0.42	0.676	-0.4508455	0.2940529
zpercept						
L1.	0.2265257	0.1540254	1.47	0.145	-0.0802426	0.5332939
L2.	-0.0901674	0.1258408	-0.72	0.476	-0.3408012	0.1604663
L3.	0.1854286	0.1238826	1.5	0.139	-0.0613049	0.4321622
L4.	0.3647755	0.1187094	3.07	0.003	0.1283451	0.6012058
L5.	-0.0420215	0.1045724	-0.4	0.689	-0.2502955	0.1662525
zbio						
L1.	-0.0168174	0.0963288	-0.17	0.862	-0.2086729	0.1750381
L2.	0.0502573	0.0773553	0.65	0.518	-0.1038091	0.2043237
L3.	-0.0342887	0.0514414	-0.67	0.507	-0.1367431	0.0681657
L4.	-0.1091203	0.0518148	-2.11	0.039	-0.2123184	-0.0059222
L5.	0.0564523	0.0542936	1.04	0.302	-0.0516828	0.1645874



According to aforementioned result, we further plot the impulse response function between the number of retweet and social process, bio process, cognitive process, perceptual process, and the number of retweet itself. Impulse response analysis represents the impact on endogenous variables when the system is subjected to external shocks. The result can be shown in figure 3.



**Fig. 3.** Impulse response analysis.

The solid lines in Figure 3 respectively show the changes of the impulse response of affect process words, bio process words, cognitive process words, perceptual process words, social process words and retweeting behavior with the increase of the prediction period. The dashed lines represent the confidence intervals of adding or subtracting two standard deviations on both sides of the corresponding impulse response images. It can be seen from impulse response analysis that cognitive process has nothing to do with the retweeting behavior. On the one hand, this may be because there is not necessarily a simple linear relationship between these variables. On the other hand, according to the cognitive evaluation theory of emotion, cognitive evaluation often affects the behavior through emotion as the intermediary. Therefore, in the Tianjin Port explosion, the influence of cognitive process on forwarding behavior may not be direct, but may be generated through emotional process.

Here, affect process fluctuate significantly in stage 1, then rise to a peak in stage 4, and gradually level off after stage 7. This may be due to the fact that emotions, especially negative emotions, can be transmitted to each other in social emergencies. After the occurrence of emergencies, negative emotions break out instantly and form a scale effect soon, thus affecting online public opinion. Therefore, emotional process words have a significant impact on retweeting behavior.

Both bio process words and social process words are relatively gentle in the initial stage, but reach their peak in the fourth stage, and then gradually tend to be flat. It can be seen that people's physiological and social processes in social emergencies are not as direct as their emotional processes, and they often exert their influence slowly in

social interaction. Therefore, the influence on retweeting behavior is also triggered after a certain period of time.

At the same time, it can be seen that the lag order of the perceptual process increases when the lag order is 1, then decreases to the lowest when the lag order is 6, and finally tends to be stable. This may be due to the fact that in the Tianjin Port explosion incident, Weibo users perceived the danger of the emergency and out of sympathy for the victims, so that the incident was quickly retweeted on Weibo and aroused to the highest point in a short period of time. However, with the rapid response of the country to the incident, the online public opinion was effectively controlled and gradually leveled off. Therefore, perceptual process words have short-term positive effects on retweeting behavior.

## 5 Conclusion

Taking the Tianjin Port explosion incident as an example, this paper constructs a VAR model to analyze the information dissemination behavior of social media users. Granger causality test results show that emotional process, perceptual process, social process and physiological process can effectively predict the information transmission behavior of users on social media, while cognitive process has no direct impact on the communication behavior. This study has three main contributions: First, through the study of the psychological process of social media users, this paper identifies the psycholinguistic factors that affect the behavior of information transmission. Secondly, through the construction of VAR model, it is found that emotional process, perceptual process, social process and physiological process are the most important factors affecting information transmission behavior, which expands the application scenario of psycholinguistics. Thirdly, by analyzing the fluctuation of online public opinion in social emergencies, this study can guide the government or emergency management departments to do a good job in public opinion prevention and response. Therefore, this study has strong practical significance.

This paper also has some limitations. First of all, this paper takes the Tianjin Port explosion as an example to predict the information transmission behavior of social media users, but whether the results are applicable in other social emergencies remains to be discussed. Second, psychological processes are composed of more than 80 types of words, and through the above literature analysis, we selected the five most important categories in the VAR model. In future studies, we should subdivide subcategories and other functional words, and analyze the predictive role of other psycholinguistic variables on the information transmission behavior of social media users.

## Acknowledgments

This work is the research results of Liaoning Economic and Social Development Research project in 2024 by the Provincial Social Science Federation (grant number: 2024lslybkt-121).

## References

1. M. Kaigo, Social media usage during disasters and social capital: Twitter and the Great East Japan earthquake. *Keio Commun. Rev.* 2012, 34, 19–35.
2. J. Yu, C. Liu, Empirical study on Micro-blog user's behavior patterns of information distribution[J]. *Modern Information*, 2015, 10, 49–54.
3. X. Z. Xu, Y. Y. Lan, X. J. Wang, Research on the influencing factors of Weibo user forwarding behavior in response to emergencies[J]. *Information Science*, 2022, 40(5): 11-18.
4. X. M. Tang, S. Q. Lai, J. Y. Liu, A literature review on social media user's information retweet[J]. *Journal of Intelligence*, 2022, 41(7): 131-137.
5. K. Zhang, X. J. Cha, Y. L. Yan, Grounded detection on influencing factors for herd forwarding behavior of health information: The case of Sina Weibo[J]. *Library and Information*, 2022(2): 10-18.
6. J. Zhou, W. M. Pan, H. J. Zhang, Key audience group identification based on emotional tendency analysis[J]. *Modern Electronics Technique*, 2021, 44(17): 31-36.
7. S. C. Ding, Z. Bao, X. Y. Liu, Research on user behavior prediction of emergency public opinion communication[J]. *Journal of Modern Information*, 2023, 43(9): 111-123.
8. X. D. Feng, J. Ma, G. Y. Jiang, Social trust, rational behavior and government Microblogs' dissemination: An empirical analysis based on text mining[J]. *Journal of the China Society for Scientific and Technical Information*, 2019, 38(09): 954-965.
9. L. Zhang, H. Wang, S. Y. Yi, Research on information acquisition effect of social media users under information epidemic[J]. *Information Research*, 2024, 315(1): 28-36.
10. N. Y. Liu, Multiple attribute decision-making method of intuitionistic linguistic considering psychological behavior of decision-maker[J]. *Operations Research and Management Science*, 2019, 28(07):11-16.
11. S. H. Bai, Individual identification model of psychological crisis based on big data and machine learning[J]. *Electronic Design Engineering*, 2023, 31(13): 17-21.
12. Pennebaker J W, Stone L D, Words of wisdom: Language use over the life span.[J]. *Journal of Personality and Social Psychology*, 2003, 85(2): 291-301.
13. Kramer A, Rodden K, Word usage and posting behaviors: modeling blogs with unobtrusive data collection methods[C]. *ACM*, 2008.
14. X. Y. Zhang. LIWC: A text analysis tool based on lexicometrics[J]. *Journal of Southwest University for Nationalities*, 2015(04): 101-104.
15. P. L. Zhu, Y. X. Xia, D. X. Jin, Dynamic mechanism of network public opinion negative emotion guidance from the perspective of emotional infection[J]. *Information Research*, 2023, 305(03): 16-23.
16. L. P. Zhu, X. Chen, Negative social emotions and governance in major public health and safety incidents [J]. *Journal of Chongqing University(Social Science Edition)*, 2023, 29(06): 295-306.
17. Z. X. Zhang, Z. Y. Wu, Media emotional contagion and analyst optimism bias: empirical evidence based on machine learning text analysis methods[J]. *Journal of Management World*, 2021, 37(01): 170-185+11+20-22.

**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.

