

Public Perception of the COVID-19 Pandemic on Chinese Social Networking Service (Weibo): Sentiment Analysis and Fuzzy-C-Means Model

Feng Han

Hokkaido University: Hokkaido Daigaku <https://orcid.org/0000-0002-6664-0516>

ZiHeng Zhang

Hokkaido University: Hokkaido Daigaku

HongJian Zhang

Hokkaido University: Hokkaido Daigaku

Tomohiko Aoki

Hokkaido University: Hokkaido Daigaku

Katsuhiko Ogasawara (✉ oga@hs.hokudai.ac.jp)

Hokkaido University <https://orcid.org/0000-0001-5474-7861>

Research

Keywords: Natural language processing (NLP), Sentiment analysis, Fuzzy c-means, COVID-19

Posted Date: June 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-592889/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Social media analysis tools have been used to monitor public sentiment and communication methods during public health emergencies.

Public health emergencies are required to better understand the impact of the crisis on the public and to provide reference material for the prevention of future public health emergencies. We are concentrating on the sentiments around the public health emergency created by COVID-19.

Objective

This study aims to better understand the impact of public health emergencies on citizens and provide reference material for future public health emergency prevention.

Methods

The Fuzzy-c-means method was used to divide the 850,083 content of Weibo from January 24, 2020, to March 31, 2020, into seven categories of emotions: fear, happiness, disgust, surprise, sadness, anger, and good. The changes in emotion were tracked over time.

Results

The results indicated that people showed "surprise" overall (55.89%); however with time, the "surprise" decreased. As the knowledge regarding the coronavirus disease 2019 (COVID-19) increased (contents about COVID-19 knowledge: from 21.16% to 4.19%), the "surprise" of the citizens decreased (from 59.95% to 46.58%). Citizens' feelings of "fear" and "good" increased as the number of deaths associated with COVID-19 increased ("fear": from 15.42% to 20.95% "good": 10.31% to 18.89%). As the infection was suppressed, the feelings of "fear" and "good" diminished ("fear": from 20.95% to 15.79% "good": from 18.89% to 8.46%).

Conclusions

In this study, the emotions and changes in emotions of Weibo users were analyzed in chronological order. The results of this study can prepare for future public health emergencies.

Background

The coronavirus disease (COVID-19) pandemic has spread in more than 200 countries and has caused many deaths [1]. As of March 25, 2021, the death toll reached 2,758,877 [2]. With the World Health Organization (WHO) pandemic statement and government action on the disease, various sentiments regarding COVID-19 have spread across the world. Over the last decade, social media analysis tools have

been used to monitor public sentiment and communication methods for public health emergencies such as the Ebola and Zika epidemics.

There are two directions in the emotion analysis area: sentiment strength detection and detection of multiple emotions. Emotional strength detection tends to classify emotions into three types: positive, neutral, and negative. Conversely, detection of multiple emotions classifies human emotions into various types.

Many studies have analyzed sentiment strength detection and detection of multiple emotions on social media, such as tweets [3, 4, 5, 6]. However, most of China's social media (Weibo) analyses perform sentiment strength detection [3, 7] and the emotions fall into only three categories: positive, neutral, and negative. In this study, seven emotions were set, and emotion classification was studied according to these seven emotions.

There are two existing sentiment analysis methods: supervised learning and unsupervised learning. Neural networks and machine learning techniques are used with supervised learning. First, the model is trained, and then it is used to classify emotions. The disadvantage of this method is that the creation of training data is labor intensive. Unsupervised learning primarily creates rules and dictionaries, which are then used to classify emotions. The disadvantage of this method is that words that are not in the rules or dictionaries cannot be analyzed. However, it is impossible to put every word in the dictionary.

The fuzzy c-means (FCM) method used in this study is an unsupervised soft computing technology. It was developed by Dunn [8] in 1973 and improved by Bezdek [9] in 1981. The soft clustering method, as compared to the hard clustering method uses a fuzzy set [10], which can better solve the problem of text ambiguity. Membership in fuzzy sets, indicates the degree of matching between the element and the set, with membership values ranging from 0 to 1. The concept of membership was extended using FCM. In this method, the membership matrix represents the membership value of the elements in multiple clusters. FCM is one of the most commonly used methods for solving fuzzy problems. Compared with other clustering methods, it is flexible and can accurately represent the degree of data affiliation [11].

The advantages of this method are as follows: 1. the number of words that constitute an emotion dictionary can be reduced; 2. it can analyze words that are not in an emotion dictionary; 3. it is suitable for accurately judging the ambiguity of people's emotions.

This study aims to better understand the impact of public health emergencies on citizens and provide reference material for future public health emergency prevention. FCM was used to analyze seven different emotions related to Weibo's content and track changes in these emotions over time.

Methods

The overall structure of the proposed method is shown in (Fig. 1). It contains five parts. Preprocessing, feature extraction, clustering, and emotion classification. All calculation methods used in this experiment were implemented in Python.

First, the raw data needs to be preprocessed. A skip-gram is then used to extract word features and convert them into computer-processable data. Next, the processed data is clustered by the FCM method. Finally, the data are classified into seven types of emotions using the clustered values of the words in each sentence.

Data set

As shown in (Fig. 2), the data source used in this study was Weibo. The collection time was from January 24, 2020 to March 31, 2020. The keywords searched were “COVID-19 outbreak status” and “COVID-19 pneumonia” and the data collected totaled 1,367,842 user contents.

Preprocessing

The input dataset is preprocessed using normalization and python code. The preprocessing tasks are as follows:

- Excluding Weibo contents that have no meaning: URLs, images, etc.
- Removing special characters: remove all special characters (punctuation marks, question marks, exclamation marks, etc.) and replace them with spaces.
- Morphological analysis: extract only nouns, adjectives, verbs, and adverbs (NAVA).

Finally, there were 850,083 contents (only NAVA) that were used as data.

Emotional dictionary

In this study, seven emotions were set based on Ekman, P, and Xu. The first six sentiments were set based on Ekman’s basic emotions [12], whereas the last was set based on Chinese local emotions [13]. (Table 1) lists the seven emotions and some of the corresponding words that represent them.

Table 1
Seven types of emotions and their representative words

emotions	representative words
fear(□)	☹️(Panic)☹️(horror)☹️(fear)☹️(fear) ☹️(worry) ☹️(horror) ☹️(scary) ☹️(shuddering) ☹️(frightened) ☹️(frightened) ☹️(fearful)...
happiness(□)	😊(Pleasant)😊(joyous) 😊(cheerful) 😊(wishful) 😊(happy) 😊(entertaining) 😊(fortunate) 😊(happy) 😊(happy) 😊(cheerful) 😊(comfortable) 😊(happy) 😊(haha)...
disgust(□)	😡(Disgust)😡(abomination) 😡(traitor) 😡(bullying) 😡(extortion) 😡(humiliation) 😡(contempt) 😡(betrayal) 😡(treacherous) 😡(wicked) 😡(treacherous) 😡(deceitful) ...
surprise(□)	😮(Wrong)😮(shocked) 😮(stunned) 😮(stunned) 😮(amazed) 😮(shocking) 😮(surprised) 😮(strangely surprised) 😮(amazing) 😮(stunned) 😮(shocked) 😮(uproar) ...
sadness(□)	😞(Grief)😞(sorrow) 😞(sadness) 😞(weeping) 😞(weeping) 😞(crying) 😞(crying) 😞(bitterness) 😞(sorrow) 😞(worry) 😞(pity) 😞(sorrow) 😞(suffering) ...
good(□)	😊(Sweetness)😊(good sentences)😊(good fun) 😊(universal love) 😊(perseverance) 😊(simple and elegant) 😊(good taste) 😊(building a country and a nation) 😊(reverence) 😊(praise) 😊(praise) 😊(praise) ...
anger(□)	😡(Rage)😡(anger) 😡(anger) 😡(grief) 😡(anger) 😡(anger) 😡(anger) 😡(anger) 😡(frustration) 😡(anger) 😡(anger) ...

Feature extraction

Feature extraction must be used to convert words in natural languages into computer-processable word vectors. In this study, the word2vec skip-gram [14] was used to extract features from the collected data.

The dimension of the word vector was set to 100 [15]. Each word was represented as a 1 × 100 vector.

Fuzzy-C-Means

The FCM method was used to cluster representative words of seven types of emotion dictionaries. The coordinates of seven centers (100 dimensions) were obtained.

Emotion classification

The membership value for each word of the seven emotions was calculated using the word vector for each word and the center coordinates of the seven emotions. The word membership value was then used to calculate the average membership value for each emotion in the sentence. The emotion with the highest degree of membership was the final emotion.

Result

Overall result

Judging from the overall results, “surprise” was the most common (475,102 cases, 55.89%). The second was “fear” (156,526 cases, 18.41%), and the third was “good” (117,799 cases, 13.86%). This is shown in (Fig. 3) and (Fig. 4).

Time series analysis

As shown in (Fig. 5), the time series of emotions can be divided into three stages: the start stage (January 24–February 6, 2020), occurrence stage (February 7–March 10, 2020), and end stage (March 11–March 30, 2020).

As shown in (Table 2), “surprise” is high at the start stage (average: 59.95%) and end stage (average: 66.17%), and low at the stage of occurrence (average: 46.58%). “Fear” is low at the start (average: 15.42%) and end (average: 15.79%), and increases at the onset (average: 20.95%). “Good” was low at the start stage (average: 10.31%) and end stage (average: 8.46%), and was high at the stage of occurrence (average: 18.89%). “Happiness” is high at the start stage (average: 12.73%) and at the stage of occurrence (average: 11.55%), and decreases at the end stage (average: 7.41%).

Table 2
Mean value of each emotion depending on the stage.

emotions	Start stage	Occurrence stage	End stage
fear(□)	15.42%	20.95%	15.79%
happiness(□)	12.73%	11.55%	7.41%
disgust(□)	0.09%	0.32%	0.53%
surprise(□)	59.95%	46.58%	66.17%
sadness(□)	1.35%	1.62%	1.38%
good(□)	10.31%	18.89%	8.46%
anger(□)	0.14%	0.09%	0.26%

Discussion

In this study, Weibo user emotions were analyzed using COVID-19-related Weibo data from January 24 to March 30, 2020; changes in Weibo user emotions were analyzed in chronological order. Specifically, Weibo users’ emotions were classified into seven categories.

Overall result

From the results of the overall analysis, it was judged that the citizens' feelings toward COVID-19 were mainly "surprised" (55.88%). This is consistent with the results of Wang et al. [16] (surprise: 53.3%). Fan et al. [17] showed that Weibo's emotions were primarily angry. Comparing the results of this study with those

of Fan et al.'s study [17, 19], it was found that the emotions of Weibo people during the outbreak of COVID-19 changed significantly from "anger" to "surprise."

Time series analysis results

Citizens' emotions regarding COVID-19 can be classified into three parts: the start stage of COVID-19, the occurrence stage of COVID-19, and the end stage of COVID-19. The stages and emotions toward it were the same as that found by previous studies on Weibo [18, 19, 20]. Compared to a previous study on Tweet [21], the emotions of the three stages and each stage are primarily the same, but the time partition of each stage is different because the user range is different. The details are as follows.

At the start of COVID-19 (January 24–February 6, 2020), people's main feelings were "surprise." The reason is that the citizens did not understand COVID-19 (contents about COVID-19 knowledge: 21.16% on the January 24, 2020) [22, 23]. At the same time, the mood of "happiness" fluctuated during this period, which was because of China's "Chinese New Year [23]."

During the COVID-19 outbreak (February 7–March 10, 2020), citizens' feeling of "surprise" decreased (from 59.95–46.58%). This is because of the large-scale dissemination of knowledge regarding COVID-19 organized by scholars, such as Zhong Nanshan. A decrease in content related to COVID-19 knowledge (number of contents for COVID-19 knowledge: 4.19% on the February 6, 2020) indicates a decrease in public interest in COVID-19 knowledge. In addition, with the increase in the number of deaths [24], the feeling of "fear" increased from 15.42–20.95%. The feeling of "good" increased from 10.31–18.89%. Because people hope the situation of the COVID-19 epidemic will get better

At the end of phase I of the COVID-19 pandemic (March 11– March 31, 2020), the feeling of "fear" decreased from 20.95–15.79% because the infectious disease was suppressed. The feeling of "good" decreased from 18.89–8.46%. However, the increase in "surprise" was inconsistent with the survey [17]. This is because the public's focus was on the epidemic situation in foreign countries such as Italy, India, Brazil, and France [25].

Limitations

This study has some limitations. First, it uses only Sina Weibo as a social media platform, and the data source for it was narrow. Because China is advancing into an aging society and most Internet users are young people, Sina Weibo users are mainly in the 18–41 age group [26]. Representing the sentiment of the population of all age groups in China is impossible.

Second, only texts were used for this study to analyze emotions. Pictograms or symbols contained in sentences were not analyzed. Pictograms and symbols contain considerable emotional information [27], and emotions are lost if not processed.

Conclusion

In this study, the FCM method was used to analyze Weibo content from January 24, 2020 to March 31, 2020. In addition, people's feelings were analyzed regarding COVID-19 pandemic in three stages over time.

Throughout the period, the public's attitude toward emergencies was a "large surprise." In the beginning, people's emotions were primarily "surprised;" however after the outbreak, people's "surprise" decreased with increasing knowledge. In addition, as the number of deaths increased, people felt "fear" and "good." At the end of phase I of the COVID-19 pandemic, people's "fear" and "good" feelings were diminished as the epidemic was suppressed. People's interest shifted from China to other countries and their concern about the situation in other countries.

From the results, it is possible to understand whether a public health emergency is a public sentiment or an idea. Our findings facilitate an understanding of public discussions and emotions about the COVID-19 pandemic among Weibo users between January 24 and March 31, 2020. By analyzing these emotions, we can provide reference materials and enable better preparation for a future public health emergency.

Abbreviations

COVID-19

Coronavirus disease 2019

NLP

Natural language processing

WHO

World Health Organization

URL

Uniform Resource Locator

FCM

Fuzzy C-Means

NAVA

nouns, adjectives, verbs, and adverbs

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding.

Authors' contributions

The first author (HF) analyzed the data, and was a major contributor in writing the (ZHJ) collected the data. All authors participated in planning the study and read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

Graduate School of Health Sciences, Hokkaido University, Japan

Feng Han, ZiHeng Zhang, HongJian Zhang, Tomohiko Aoki, Katsuhiko Ogasawara

Corresponding author

Correspondence to Ogasawara Katsuhiko.

References

1. World Health Organization. 2020.05.18. Coronavirus disease 2019 (COVID-19) Situation. Report. 83 www.who.int/docs/default-source/coronaviruse/situation-reports/20200412-sitrep-83-covid-19.pdf?sfvrsn=697ce98d_4(April 30, 2021)
2. The report of Coronavirus disease 2019[in Chinese]wp.m.163.com/163/page/news/virus_report/index.html?_nw_=1&_anw_=1.(March 25, 2021)
3. Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, Hon SD, Park J, Dang P, Lipsky MS 2020 Aug 18. Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence. Journal of Medical Internet Research, 22(8), e22590. doi: 10.2196/22590. PMID: 32750001; PMCID: PMC7438102.
4. Dyer J, Kolic B 2020. Public Risk Perception and Emotion on Twitter During the Covid-19 Pandemic. Applied Network Science, 5(1), 99. doi: 10.1007/s41109-020-00334-7. PMID: 33344760; PMCID: PMC7739810.

5. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, Yang Y 2020 May 22. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. *JMIR Public Health and Surveillance*, 6(2), e19447. doi: 10.2196/19447. PMID: 32412418; PMCID: PMC7247466.
6. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, Zhu T 2020 Nov 25. Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. *Journal of Medical Internet Research*, 22(11), e20550. doi: 10.2196/20550. PMID: 33119535; PMCID: PMC7690968.
7. Li S, Wang Y, Xue J, Zhao N, Zhu T 2020 Mar 19. The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. *International Journal of Environmental Research and Public Health*, 17(6), 2032. doi: 10.3390/ijerph17062032. PMID: 32204411; PMCID: PMC7143846.
8. Dunn JC 1973. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3), 32.
9. Bezdek JC, Ehrlich R, Full W 1984. FCM: The Fuzzy c-means Clustering Algorithm. *Computers and Geosciences*, 10(2-3), 191.
10. Zadeh LA 1965. Fuzzy Sets. *Information and Control*, 8(3), 338.
11. Murthy C. 2014, Mod-06 Lec-41 FCM and soft-computing techniques.
12. Ekman P. 1999a. Basic Emotions. In: *The handbook of cognition and emotion*, ed. Dalglish, T. & Power, T. pp. 45-60. Wiley.
13. Xu Linhong, etc. "The Construction of Emotional Vocabulary Ontology." *Journal of Information* 27.2(2008):180-185. [in Chinese]
14. Mikolov T, Corrado G, Chen K, Dean J. 2013. Efficient Estimation of Word Representations in Vector Space, *Proceedings of the International Conference on Learning Representations (ICLR 2013)*. pp. 1-12.
15. Wang Y, Liu S, Afzal N, Rastegar-Mojarad M, Wang L, Shen F, Kingsbury P, Liu H 2018. A Comparison of Word Embeddings for the Biomedical Natural Language Processing. *Journal of Biomedical Informatics*, 87, 12. doi.org/10.1016/j.jbi.2018.09.008.
16. Wang Yi, Gao Junling, Chen Hao, et al. Public media exposure during the new crown pneumonia epidemic and its relationship with mental health. *Fudan Journal (Medical Edition)* 2020, 6. [in Chinese]
17. Fan R, Zhao J, Chen Y, Xu K 2014. Anger Is More Influential Than Joy: Sentiment Correlation in Weibo. *PLOS ONE*, 9(10), e110184. doi: 10.1371/journal.pone.0110184 dx.plos.org/10.1371/journal.pone.0110184.
18. Chen Xingshu, Chang Tianyou, Wang Haizhou, et al. Spatiotemporal analysis of public opinion evolution of the "new crown pneumonia epidemic" based on Weibo data. *Journal of Sichuan University (Natural Science Edition)* 2020, 2. [in Chinese]
19. Yu S, Eisenman D, Han Z 2021. Temporal Dynamics of Public Emotions During the COVID-19 Pandemic at the Epicenter of the Outbreak: Sentiment Analysis of Weibo Posts From Wuhan. *Journal of Medical Internet Research*, 23(3), e27078. doi.org/10.2196/27078.

20. Zhao Y, Cheng S, Yu X, Xu H 2020. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. *Journal of Medical Internet Research*, 22(5), e18825. doi.org/10.2196/18825.
21. Boon-Itt S, Skunkan Y 2020. Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. *JMIR Public Health and Surveillance*, 6(4), e21978. doi.org/10.2196/21978.
22. Chen H, Zhang K 2020 Nov 01. Insight Into the Psychological Problems on the Epidemic of COVID-19 in China by Online Searching Behaviors. *Journal of Affective Disorders*, 276, 1093. doi: 10.1016/j.jad.2020.07.128.
23. Zhu B, Zheng X, Liu H, Li J, Wang P 2020. Analysis of Spatiotemporal Characteristics of Big Data on Social Media Sentiment With COVID-19 Epidemic Topics. *Chaos, Solitons, and Fractals*, 140, 110123. doi.org/10.1016/j.chaos.2020.110123.
24. China CDC. (CCDC) weekly.chinacdc.cn/news/TrackingtheEpidemic.htm.
25. Wang J, Zhou Y, Zhang W, Evans R, Zhu C 2020. Concerns Expressed by Chinese Social Media Users During the COVID-19 Pandemic: Content Analysis of Sina Weibo Microblogging Data. *Journal of Medical Internet Research*, 22(11), e22152. doi.org/10.2196/22152.
26. Sina microblog data center. 2020.02.21. 2018. Sina microblog user development report [in Chinese] data.weibo.com/report/reportDetail?id=433. (March 25, 2021)
27. Aldunate N, González-Ibáñez R 2016. An Integrated Review of Emoticons in Computer-Mediated Communication. *Frontiers in Psychology*, 7, 2061. doi.org/10.3389/fpsyg.2016.02061.

Figures

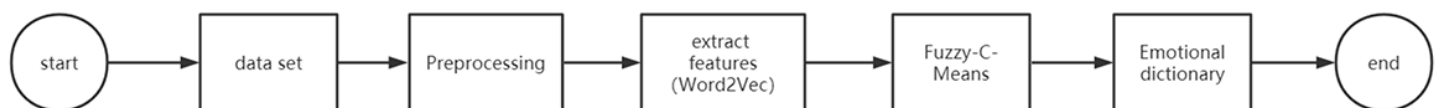


Figure 1

Overall structure of the proposed method

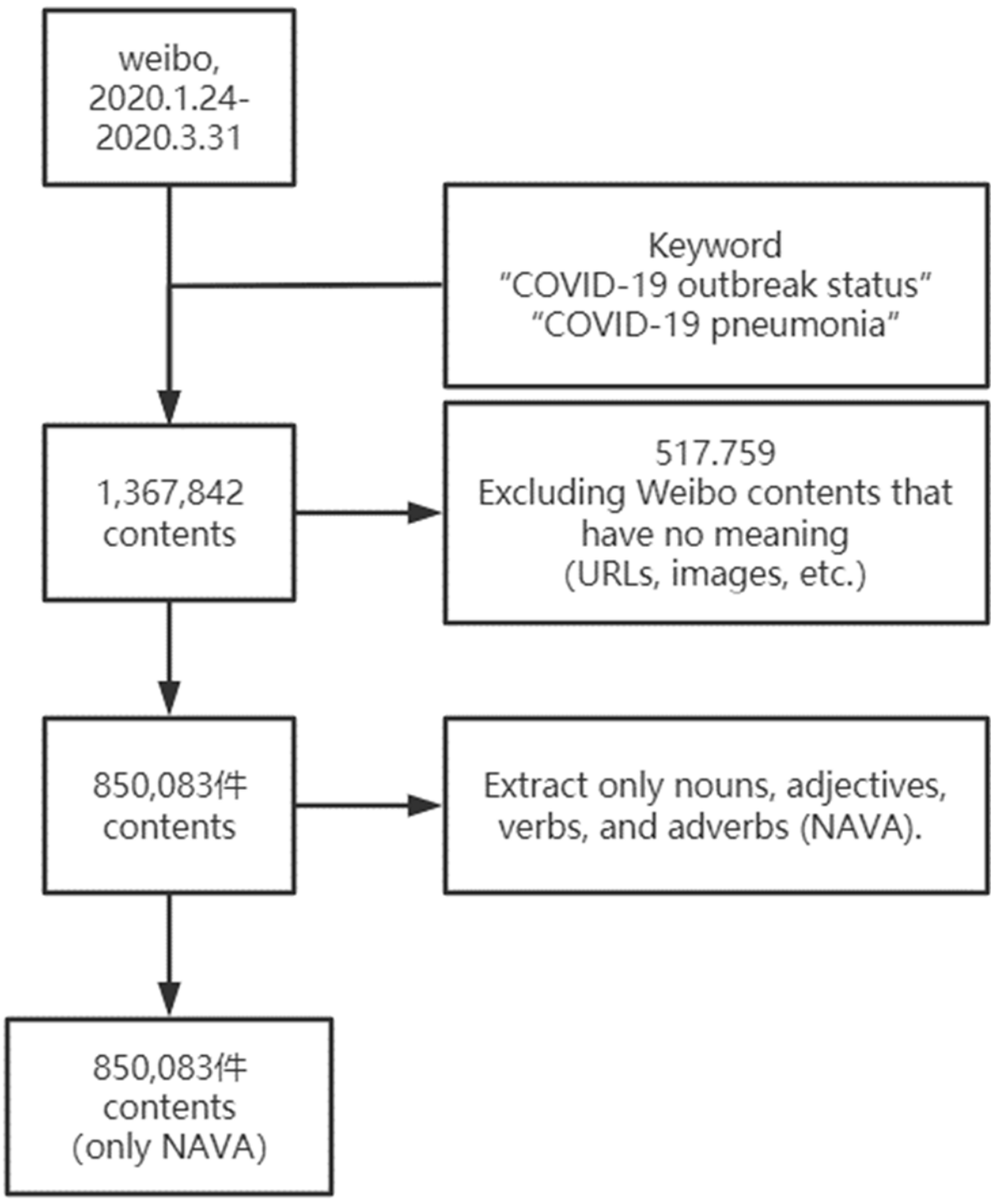


Figure 2

Data and preprocessing

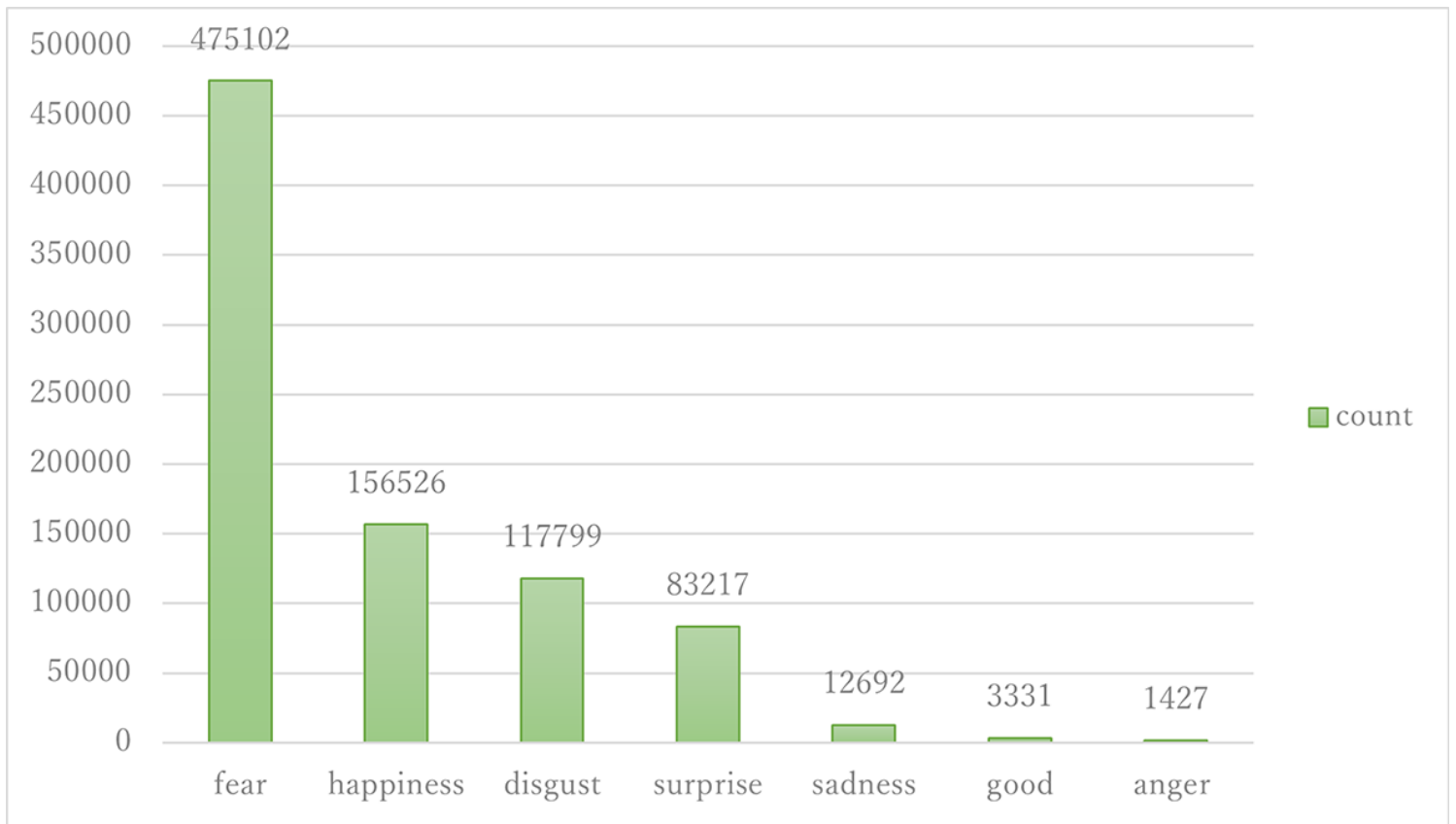


Figure 3

Overall results (quantity)

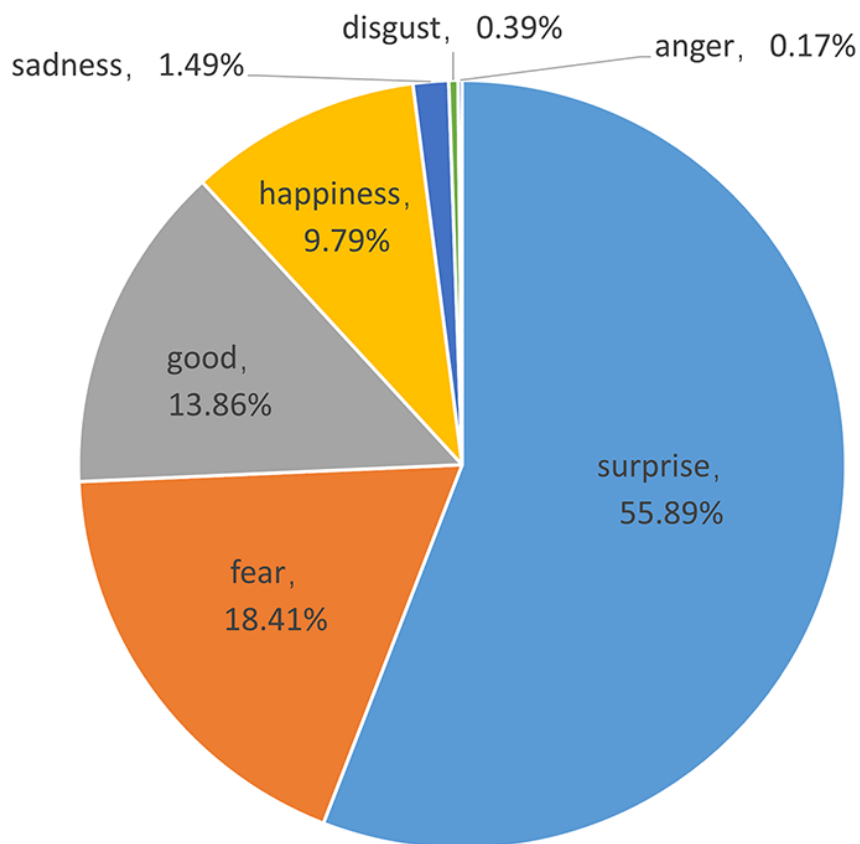


Figure 4

Overall results (ratio)

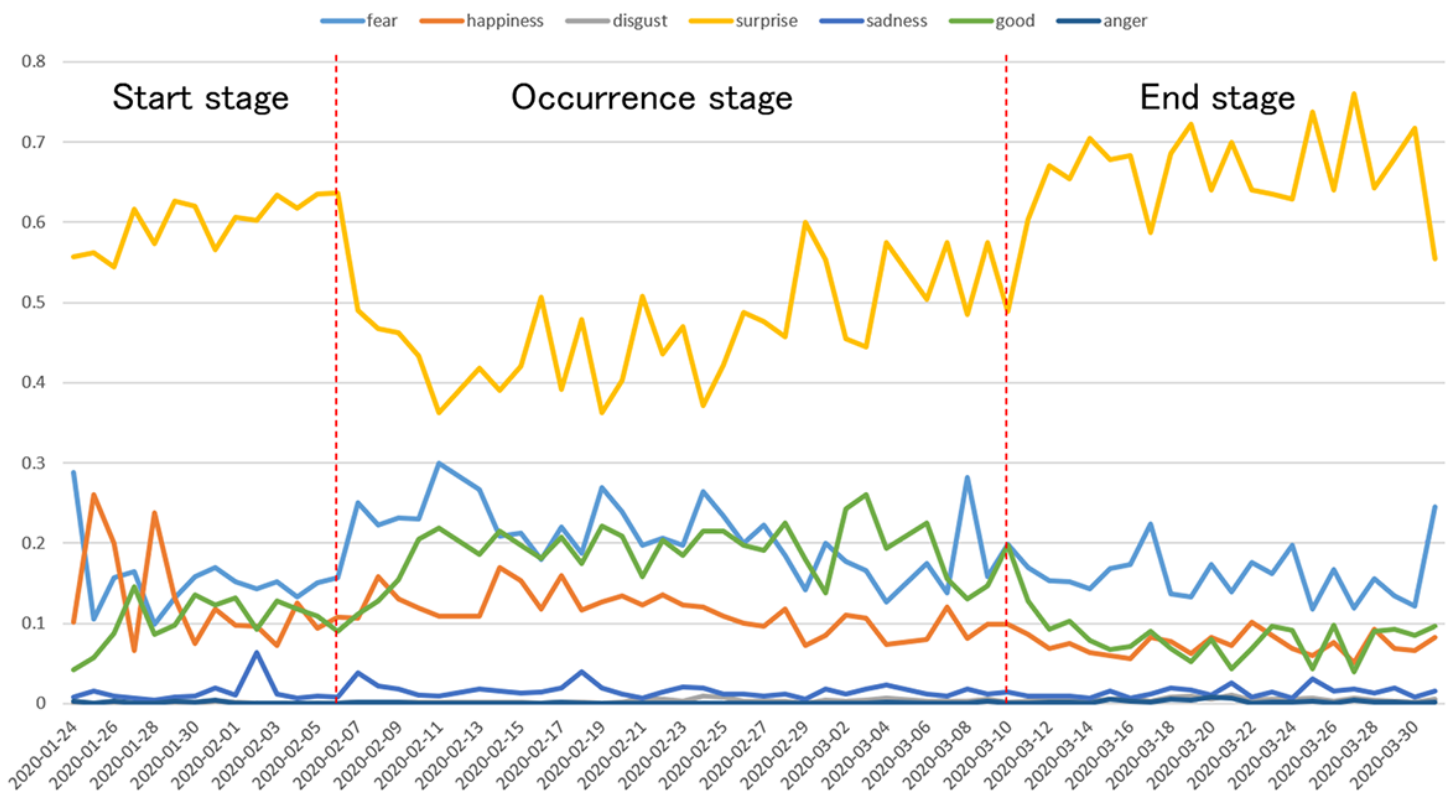


Figure 5

Results of time series analysis.