

# A Sentiment Analysis Approach to Discover Public Panic: Based on Weibo Covid-19 Data

Wanjun Wu

Bytedance Data Analysis Group, Beijing, China

Email: wuwanjun.jennifer@bytedance.com

**How to cite this paper:** Wu, W.J. (2022) A Sentiment Analysis Approach to Discover Public Panic: Based on Weibo Covid-19 Data. *Social Networking*, 11, 33-39.  
<https://doi.org/10.4236/sn.2022.113003>

**Received:** July 2, 2022

**Accepted:** July 26, 2022

**Published:** July 29, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

**Background:** Weibo is a Twitter-like micro-blog platform in China where people post their real-life events as well as express their feelings in short texts. Since the outbreak of the Covid-19 pandemic, thousands of people have expressed their concerns and worries about the outbreak via Weibo, showing the existence of public panic. **Methods:** This paper comes up with a sentiment analysis approach to discover public panic. First, we used Octoparse to obtain Weibo posts about the hot topic Covid-19 Pandemic. Second, we break down those sentences into independent words and clean the data by removing stop words. Then, we use the sentiment score function that deals with negative words, adverbs, and sentiment words to get the sentiment score of each Weibo post. **Results:** We observe the distribution of sentiment scores and get the benchmark to evaluate public panic. Also, we apply the same process to test the mass sentiment under other topics to test the efficiency of the sentiment function, which shows that our function works well.

## Keywords

Sentiment Analysis, Data Analysis, Covid-19, Micro-Blogdata

---

## 1. Introduction

### 1.1. Weibo

Weibo is a Twitter-like micro-blog platform in China where people post their real-life events as well as express their feelings in short texts. Since its launch in May 2009, Weibo has been a universal real-time information network to help people discover what is happening [1]. During the Covid-19 pandemic, thousands of individuals and media organizations follow the latest news and express their concerns about the outbreak via Weibo, which gives us a chance to discover valuable intelligence from the massive user-generated text streams [2].

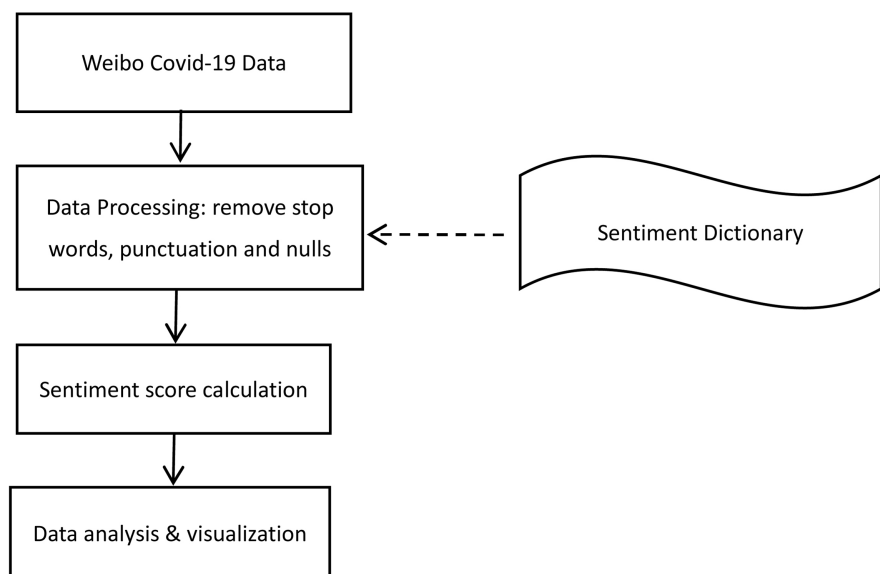
General anxiety and panic can be seen in the flood of posts related to the pandemic, and those emotions will be amplified in the development and revolutions of the viruses. If there is no timely response and handling, public fear and panic can reduce people's judgement, which may lead to spreading rumours and even vicious incidents. Therefore, timely, credible and accurate information release is critical [3].

## 1.2. Sentiment Analysis

In order to discover the public panic, we can use sentiment analysis to deal with Weibo Covid-19 data since Weibo data expose individuals' real-life status as well as what they really think. Sentiment analysis is a powerful technique to dig out people's ideas and feelings about a given text [4], either with a machine-learning or lexicon-based approach [5]. The machine-learning approach selects features of a given text, and uses Naive Bayes, Max Entropy, Support Vector Machine and other classifiers for sentiment classification [6]. The Lexicon-based approach calculates positive and negative sentiment words to classify a given text, relying on an open-source dictionary. In this paper, we choose the lexicon-based approach, which comprehensively considers a sentence's components and finally calculates the sentiment score.

## 1.3. Paper Structure

This paper comes up with a sentiment analysis approach to discover public panic, and the overall methodology framework is shown in **Figure 1**. First, we used a web scraping tool Octoparse to obtain Weibo posts about the hot topic Covid-19 Pandemic. Second, we break down those sentences into independent words and clean the data by removing stop words. Then, we use the sentiment score function that deals with negative words, adverbs, and sentiment words to



**Figure 1.** The methodology structure of the paper.

get the sentiment score of each Weibo post. Finally, we observe the distribution of sentiment scores and get the benchmark to evaluate public panic.

## 2. Materials and Methods

### 2.1. Data

In this paper, we used Octoparse to obtain Weibo posts related to the hot topic Covid-19 Pandemic in 7 May. Octoparse is a web scraping tool that provides data extraction services to grab data under a certain hot topic on Weibo.

### 2.2. Methodology

#### 2.2.1. Framework

A Chinese sentence's scientific breakdown and analysis are vital to sentiment analysis. First, we passed every post through Jieba (an open-source tool) to break the whole sentences into several independent words, like the process in **Figure 2**.

Second, we removed stop words from the split sentences. Stop words are common and high-frequency words like “a”, “the”, “of”, “and” [7], which are negligible in sentiment analysis. This process removed those unimportant words and made only those keywords left, reducing the dimension of the data that we need to handle.

The next step was dealing with the negative and adverb words. Negative words are essential when determining whether a sentence's attitude is positive or negative. For example, the sentiment of “I like watching basketball games.” and the sentiment of “I DO NOT like watching basketball games” are different. Likewise, adverb words play an important role in determining the emotional intensity of a sentence. The intensity of “I like watching basketball games very much.” is slightly stronger than the version without “very much”. The *Sentiment Score Calculation chapter* will explain the processing details of negative and adverb words.

#### 2.2.2. Sentiment Score Calculation

In order to comprehensively consider the components of a sentence as we discussed above, we conducted the sentiment score function to quantify each Weibo post:

$$\text{Sentiment Score of a post} = \left( \sum_{i,j} a_i \times (-1)^j \times w_i \right) / (\# \text{words in the sentence})$$

$i$ : the number of sentiment words in the sentence.

$w_i$ : the score of sentiment word  $i$ .

$j$ : the number of negative words in front of the sentiment word  $i$ .

$a_i$ : the score of adverb word in infront of sentiment word  $i$ .

我喜欢看篮球比赛。 → 我, 喜欢, 看, 篮球, 比赛。  
 I like watching basketball games. → I, like, watching, basketball, games.

**Figure 2.** The breakdown process of a Chinese sentence.

First, we choose Boson NLP dictionary to assign the raw score  $w_i$  of each sentiment word. Boson NLP dictionary assigns each sentiment word with an emotional intensity score. It is the most popular Chinese word segmentation method because of its context-specific lexicons such as news and social media texts [8].

Next, to deal with negative words, we count three words index above every sentiment word to check whether there are negative words. Common Chinese negative words are contained in the negative words dictionary. If one negative word exists, we will multiply  $-1$  on the raw score  $w_i$ . If  $N$  negative words exist, we would multiply minus  $(-1)^N$  on the raw score  $w_i$ . In this way, we can successfully add the impact of negative words into the sentiment score.

Likewise, when handling adverb words, we counted 3 three words index above every sentiment word to check the adverb's existence. Different adverb word owns different emotional intensity. For example, the intensity of "a little bit" is slighter than "so much". Adverb words dictionary contains universal Chinese adverbs, which are marked with different scores due to their intensity. Once we find the existence of an adverb related to the certain sentiment word  $w_i$ , we multiply the intensity score  $a_i$  to  $w_i$  to qualify the impact of the adverb.

After calculating the scores of each sentiment word of the sentence, we add all those scores together to get a final one. Then, we divide the final score by the number of the sentiment words in the sentence to normalize the impact of the length of the sentence.

### 3. Data Analysis

#### 3.1. Data Processing & Visualization

After passing Weibo Covid-19 data through the sentiment score function, we can get the score corresponding to each post. Top 2 & the bottom 2 posts are shown in the [Table 1](#) & [Table 2](#).

**Table 1.** The top 2 post's scores are calculated by the sentiment score function, and we translate the posts' content into English for ease of understanding.

Top 2 Posts Scores	Top 2 Posts Content (Translated in English by Google)
0.965	#Shanghai college entrance examination postponement# Let's take a look at the 2021 college students who are fortunate enough to be sandwiched between the two-year college entrance examination postponed due to the epidemic. I wish this year's college students great.
0.962	#JOYANG Sunshine Service#, I would like to thank the Joyoung after-sales for the Joyoung new wall breaker P771 pushed to me during the epidemic. It solved the problem in my life during the epidemic that I couldn't go out for consumption, so I could enjoy a variety of food at home with more peace of mind. The machine performance introduction will be shared with you, thanks to Joyoung@Joyoung Sunshine Service

The top & bottom scored posts indicate that our calculating function works well. The top 2 posts show the author's wishes and hopes, while the bottom two posts state people's complaints about the inconvenience brought by the pandemic.

On the one hand, the output sentiment scores follow the normal distribution (shown in **Figure 3** & **Table 3**), indicating that most posts just state the fact and do not contain too much emotional catharsis. On the other hand, we should pay more attention to those posts whose scores are beyond  $3\text{-}\sigma$ , since those posts show people's more intense emotions.

### 3.2. An Approach to Discover Public Panic

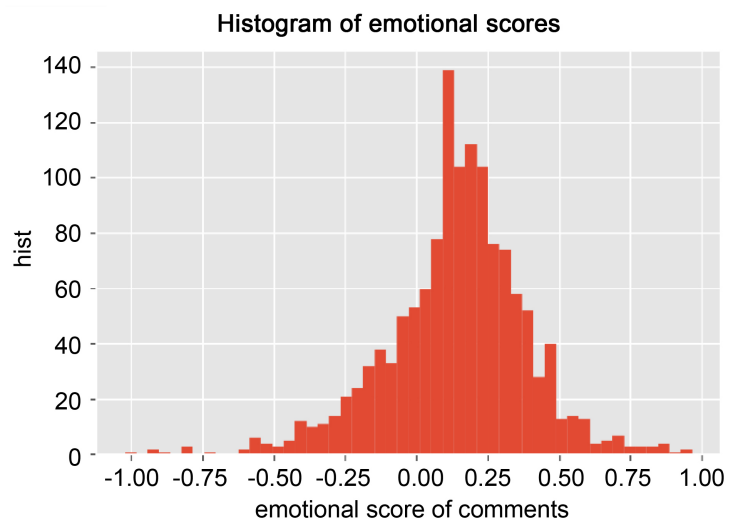
To demonstrate the efficiency of our score calculation strategy, we grab Weibo data under tag *#Come on for the college entrance examination*, where most people show their wishes and hopes for the coming exam. The distribution of

**Table 2.** The bottom 2 post's scores are calculated by the sentiment score function, and we translate the posts' content into English for ease of understanding.

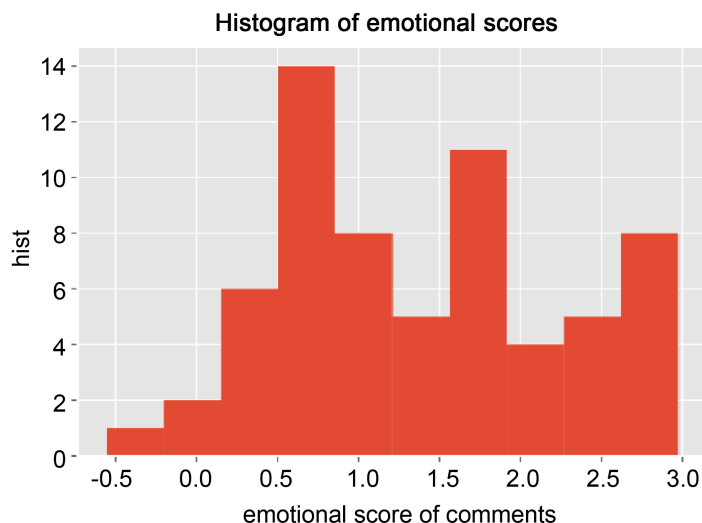
Bottom 2 Posts Score	Bottom 2 Posts Content (Translated in English by Google)
-1.02	This damn Covid-19 made me and my boyfriend become the Cowherd and Weaver Girl.
-0.93	I really hate this epidemic because the epidemic has disrupted a lot of plans.

**Table 3.** Distribution of output post scores under #Covid-19.

mean	std	min	25%	50%	75%	max
0.14	0.25	-1.02	0.01	0.15	0.29	0.96



**Figure 3.** Distribution of output post scores under #Covid-19.



**Figure 4.** Distribution of output post scores under #Come on for the college entrance examination.

**Table 4.** Distribution of output post scores under #Come on for the college entrance examination.

#Tag	mean	std	min	25%	50%	75%	max
#Come on for the college entrance examination	1.39	0.89	-0.56	0.73	1.22	1.97	2.98
#Covid-19	0.14	0.25	-1.02	0.01	0.15	0.29	0.96

sentiment scores under this tag is more favorable than those under #Covid-19, which is shown in **Figure 4** and **Table 4**, indicating that our function is efficient in distinguishing the mass sentiment under different topics.

To discover public panic, we can use the proportion of those posts whose scores are under zero as a benchmark. If that proportion is close to 50% or even greater than 50%, we should pay attention to the significant public panic under specific topics. In the #Covid-19 case, the proportion of those posts whose scores are under zero is 23.5%. Though some people express their worries and complaints about the pandemic, many still show positive attitudes towards it.

#### 4. Conclusions & Discussion

This paper presents a sentiment analysis approach to discovering public panic via Weibo data. First, we used Octoparse to obtain Weibo posts about the hot topic Covid-19 Pandemic. Second, we break down those sentences into independent words and clean the data by removing stop words. Then, we use the sentiment score function that deals with negative words, adverbs, and sentiment words to get the sentiment score of each Weibo post.

We observe the distribution of sentiment scores and get the benchmark to

evaluate public panic. Also, we apply the same process to test the mass sentiment under other topics to test the efficiency of the sentiment function, which shows that our function works well.

To further improve our method, on the one hand, we can choose different #Topics to get enough distribution data to get a confident interval of the benchmark to evaluate the public panic. On the other hand, we can improve our sentiment dictionary and adverb dictionary to get a more precise sentiment function.

## Funding Statement

This work is sponsored by Shanghai Pujiang Program (20PJ1418400).

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- [1] Bai, H. and Guang, Y. (2016) A Weibo-Based Approach to Disaster Informatics: Incidents Monitor in Post-Disaster Situation via Weibo Text Negative Sentiment Analysis. *Natural Hazards*, **83**, 1177-1196.  
<https://doi.org/10.1007/s11069-016-2370-5>
- [2] Chung, S. and Aring, D. (2018) Integrated Real-Time Big Data Stream Sentiment Analysis Service. *Journal of Data Analysis and Information Processing*, **6**, 46-66.  
<https://doi.org/10.4236/jdaip.2018.62004>
- [3] Ma, C. and Yan, X.K. (2020) Research Progress in Psychological Stress Response and Prevention and Control Strategies of COVID-19. *Journal of Jilin University (Medicine Edition)*, **46**, 649-654.
- [4] Karamitsos, I., Albarhami, S. and Apostolopoulos, C. (2019) Tweet Sentiment Analysis (TSA) for Cloud Providers Using Classification Algorithms and Latent Semantic Analysis. *Journal of Data Analysis and Information Processing*, **7**, 276-294.  
<https://doi.org/10.4236/jdaip.2019.74016>
- [5] Redhu, S., *et al.* (2018) Sentiment Analysis Using Text Mining: A Review. *International Journal on Data Science and Technology*, **4**, 49-53.  
<https://doi.org/10.11648/j.ijdst.20180402.12>
- [6] Xie, L.X., Zhou, M. and Sun, M.S. (2012) Hierarchical Structure Based Hybrid Approach to Sentiment Analysis of Chinese Micro Blog and Its Feature Extraction. *Journal of Chinese Information Processing*, **26**, 73-83.
- [7] Asghar, M.Z., *et al.* (2014) A Review of Feature Extraction in Sentiment Analysis. *Journal of Basic and Applied Scientific Research*, **4**, 181-186.
- [8] Guo, L.M., *et al.* (2019) Collaborative Filtering Recommendation Based on Trust and Emotion. *Journal of Intelligent Information Systems*, **53**, 113-135.  
<https://doi.org/10.1007/s10844-018-0517-4>