ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0059

Comparative Sentiment Analysis of Hurricane Ida and Typhoon In-fa

Yunchao Hou

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu, 214122, China

Abstract

Analysis of sentiment polarity of public opinions towards disasters has become popular in recent years for their usefulness in policy making. However, the comparative analysis of public opinions on cyclones in two nations is barely seen. It can provide a multicultural perspective for policy makers and public responses in different cultures facing disasters for social science researchers. Taking Hurricane Ida and Typhoon In-fa as instances, the research firstly collected the related posts from Twitter and Weibo and have them processed to draw a map, several wordclouds and charts. Then, the research discovered several factors contributing to the public sentiment. Lastly, the research found some interesting similarities and differences between Chinese and Americans' sentiment towards hurricanes.

Keywords

Sentiment classification; Hurricane Ida; Typhoon In-fa; Disaster-related tweets.

1. Introduction

Twitter and Weibo are two prevalent social platforms in the US and China, where people share vast amount of information publicly. They are perfect platforms for sentiment analysis for the conciseness, timeliness and rich emotions of their posts. [1] [2]

China and the United States are two of the countries most affected by the hurricane. Hurricane Ida and Typhoon In-fa alone caused $50 billion and $2 billion in damage, claiming 6 and 116 lives respectively. Hence, every time a cyclone forms, people in both countries give much attention to it, sending hundreds of thousands of posts.

Vast negative tweets have been proved to have power to lead to public panic, hatred or even incidents endangering social security. [3] Therefore, sentiment analysis of tweets during hurricanes is particularly important in order to better understand how people's emotions change and take actions to avoid mass panic.

Hurricane Ida was the second most damaging and intense hurricane to strike Louisiana, the US. It made landfall on August 29, 3 days after its formation, and dissipated on September 4. Typhoon In-fa, making landfall in Zhejiang, China on July 25, 2021, has become the second wettest tropical cyclone on record in China. It was formed on July 16 and dissipated on July 31.

This research focuses on the tweets on the topic of Hurricane Ida from August 24, 2021 to September 3, 2021 and microblog posts on Weibo about Typhoon In-fa between July 23 and July 30, 2021. The research collected relevant posts and the number of likes from both Twitter and Weibo. Additionally, for Weibo microblog posts, the research also collected the province of each post's owner and the device from which they sent the posts.

ISSN: 2637-6067
DOI: 10.6918/IJOSSER.202203_5(3).0059

2. Method

2.1. Data collection

The research used keywords "Hurricane OR Ida" to search for tweets related to Hurricane Ida posted between August 24,2021 and September 3,2021, with the help of snscrape package. [4]

Snscrape collects not only the contents of the tweets, but also their sending time, number of likes and owner's ID.

As to Typhoon In-fa, the research adopted selenium package in Python to collect the content, sending time, number of likes, user homepage URL and sending source of each microblog post. We set the search keyword as "Typhoon In-fa" and date range from July 23,2021 to July 30,2021.

2.2. Data preprocessing

2.2.1. Contents

Data collected from both platforms contained garbled codes, unwanted URLs and marks like "@", "#". We removed them using regular expressions after patterns were manually identified.

2.2.2. Sending source

The devices from which the Weibo microblog posts were sent were mixed with customized adjectives. We extracted the device name like iPhone, HUAWEI from a jumble of text by matching and replacing original text with its device name, using regular expression. For example, we replaced "My lovely purple iPhone11" with "iPhone".

For those texts that are not in our list of device names, we renamed them as "others".

2.2.3. Location

Since we had collected the homepage URLs, location were easily obtained by entering homepage of each sender.

In most cases, a location string contains not only the province but also the city or district the user sets. Therefore, we deleted the name of cities or districts by leaving just first two characters, and manually restore the province names that are three-character long in Excel.

However, the location string could be a name of a company if the account is company-owned and it could be null or a strange string. To resolve this problem, we set all null strings and those containing no location as 'others' in Excel. Then, we manually decided which province the companies were in.

2.3. Word clouds

The research used wordcloud and jieba package in Python to make word clouds to find the trend of public opinion. In order to forbid useless words from showing up, we expanded the given stopwords in wordcloud package. For example, "let", "don t" were added to stopwords before generating Ida word clouds and 'bring', 'discover' were added to stopwords before generating In-fa word clouds.

2.4. Geographic analysis

The research drew a map that shows the disparity in the number of related microblog posts by province in China using FineBI, a website for online data map generation. We divided the number of microblog posts into five categories (0-200,201-500,501-1000,1001-3000,>3000), with each represented by a different depth of blue.

2.5. Sentiment analysis

Techniques like deep learning, machine learning and lexicon based method are frequently adopted to address sentiment analysis problems. [5] BERT is a revolutionary deep learning model that caused a huge sensation for its remarkable performance in all scenarios when it was

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0059

first introduced in 2018. [6] Since it can be easily fine-tune, it is highly flexible for all types of tasks.

In this research, we used a modified BERT model called DistilBert for sentiment analysis of tweets in English. DistilBert is 60% faster and retains 97% understanding of language compared to BERT model. [7] It can divide tweets into six emotional categories, which are joy, love, surprise, anger, sadness and fear. We classify the first three of the above six sentiments as positive sentiment and the last three as negative sentiment.

For Weibo microblog posts, we took advantage of Tencent Cloud API, one of the best API that detect sentiment from Chinese sentences, to classify sentiments into positive, neutral and negative.

3. Results

3.1. Data summary

We have collected 459,917 related tweets in English from Twitter, from August 24, 2021 to September 3, 2021. 198,189 (43.1%) tweets have positive sentiment whereas 261,728 (56.9%) tweets have negative one. Joy accounts for 97.1% of positive tweets and anger, fear, sadness takes up 52.4%, 24.7% and 22.9% of negative tweets respectively.

74,018 related Chinese microblog posts were collected from Weibo, from July 23, 2021 to July 30, 2021. 16,363 (22.1%) microblog posts contain positive sentiment, 15,633 (21.1%) microblog posts contain negative one and 42,022 (56.8%) microblog posts contain neutral one.

3.2. Word clouds

3.2.1. Hurricane Ida

Broadly speaking, word clouds of Hurricane Ida demonstrated the changes of hurricane's location in that some states, cities or regions like Louisiana, New Orlean, Gulf Coast, New York were mentioned over time.

Specifically, we found that public attitudes towards the hurricane changed over time. Words showing insouciance came top and many former disastrous hurricanes like Hurricane Harvey were brought up for quips before the hurricane was upgraded to Category 1 on August 27. After that, people tended to be wary and the some of the most frequently mentioned words are like "prepare", "evacuate". Since August 29, when hurricane made landfall, empathetic words like "prayer", "thank" became ubiquitous. From August 31 onwards, some words indicating fear, anger and sadness appeared in the word clouds, like "shit", "remnant", "damage".

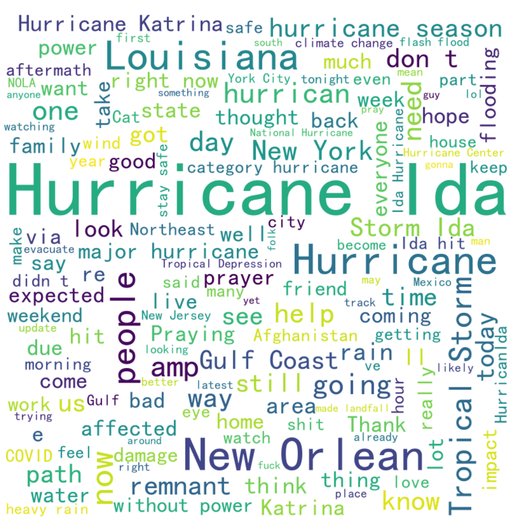

Figure 1. Word cloud of Hurricane Ida

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0059

3.2.2. Typhoon In-fa

In general, many locations were mentioned like Shanghai, Zhejiang. However, most top words were not strongly tied with emotions. Instead, they seemed to be covering the situation or reflecting the appeal. Typical words are "make landfall", "Typhoon Blue Warning", "air pressure", "prepare", "hope", "stay safe".

Concretely speaking, words calling for precautions prevailed before July 25, when the typhoon made landfall. After that, most words were describing the places hit by In-fa and its movements.

3.2.3. Comparison

It is obvious that hot words in tweets revealed more emotions than those in Weibo microblog posts while the latter reported more real-time information about the typhoon and the situation of the affected areas. Simultaneously, tweets were more colloquial, whereas microblog posts had more written language.

Both tweets and microblog posts focused on the location of hurricane and typhoon and had the voices of people calling for precautions and rescue after typhoon landfall.

3.3. Geographic analysis

As shown in Figure 1, the number of microblog posts is highly correlated with the degree of typhoon impact. Zhejiang, the province where typhoon made landfall, ranks first with the number of microblog posts reaching 11043, followed by Jiangsu (4894), Shanghai (4734), Shandong (4480). The above-mentioned provinces and municipality are the first to be hit by Typhoon In-fa. Also, other provinces and municipalities affected by the typhoon generated many microblog posts as well, such as Beijing (2628), Anhui (2609), Henan (1233). The remaining areas showed a positive correlation between number of microblog posts and population size.

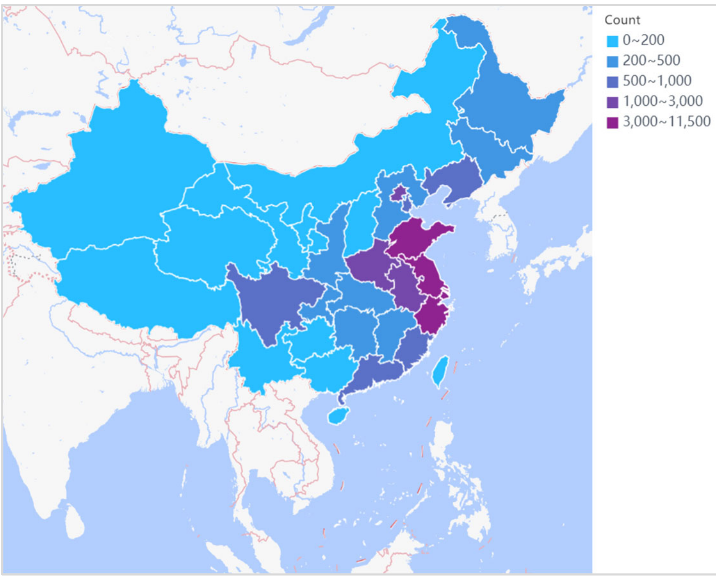

Figure 2. Map of the number of typhoon firework Weibo releases by province

3.4. Sentiment analysis

3.4.1. Hurricane Ida

Figure 2 and Figure 3 shows that positive sentiment dominated until hurricane made landfall on August 29. On August 30, the number of tweets peaked and negative tweets significantly surged. Since then, the number of tweets of both sentiments has been in a downward trend though September 2 witnessed another rise of negative tweets.

ISSN: 2637-6067
DOI: 10.6918/IJOSSER.202203_5(3).0059

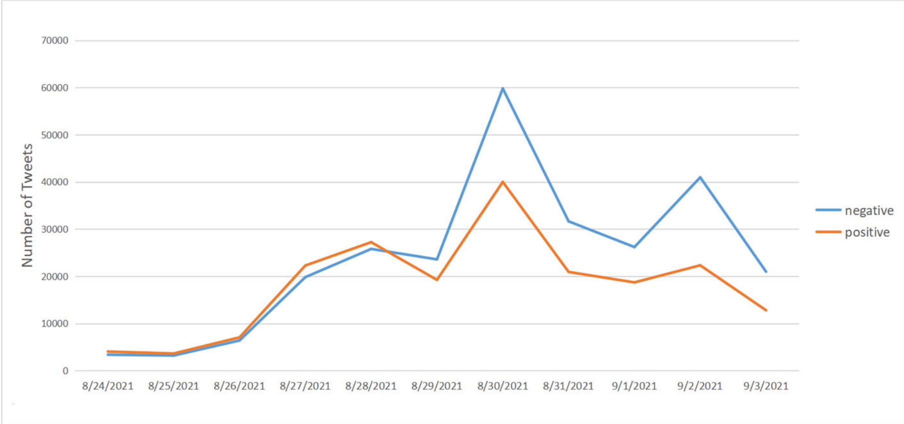

Figure 3. Number of positive and negative tweets about Hurricane Ida

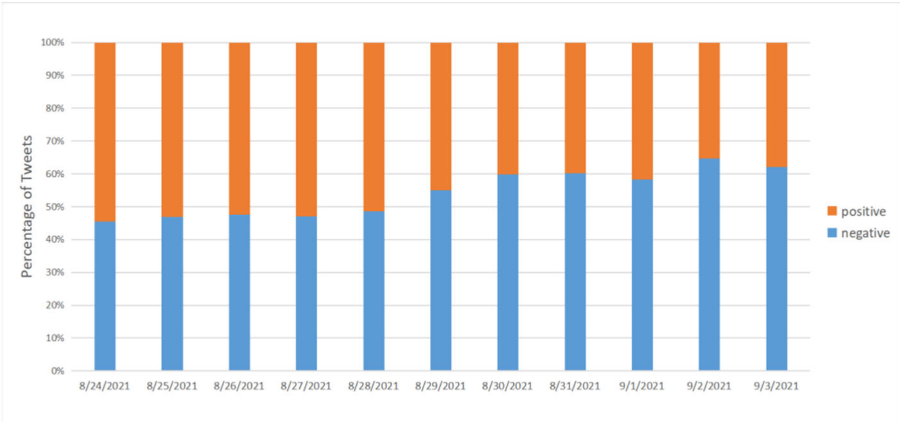

Figure 4. Proportion of positive and negative tweets about Hurricane Ida

3.4.2. Typhoon In-fa

As shown in Figure 4, microblog posts containing negative sentiment reached a plateau between July 25 and July 28, occupying the main body.

Figure 5 shows that positive sentiment took a great advantage of number before July 25 while negative posts exceeded positive ones at the night of July 24 for the first time. Additionally, a pulse of negative microblogs on the evening of July 27 can be seen. Generally speaking, the total number of microblog posts have dwindled since July 25.

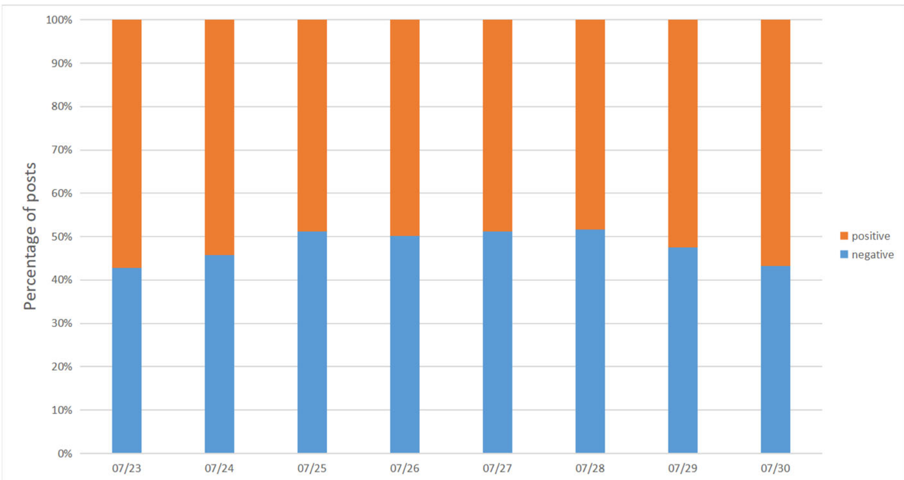

Figure 5. Proportion of positive and negative microblog posts about Typhoon In-fa

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0059

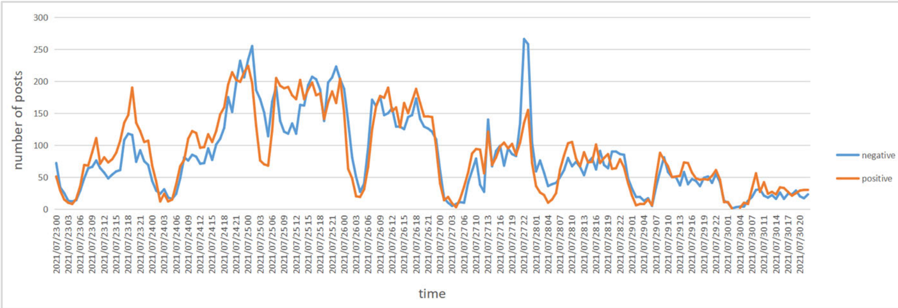

Figure 6. Number of positive and negative microblog posts about Typhoon In-fa over time

According to Figure 6, the vast majority of microblog posts were sent between 7:00 and 23:00 (87.82%). A drop in the number is seen during the midday hours.

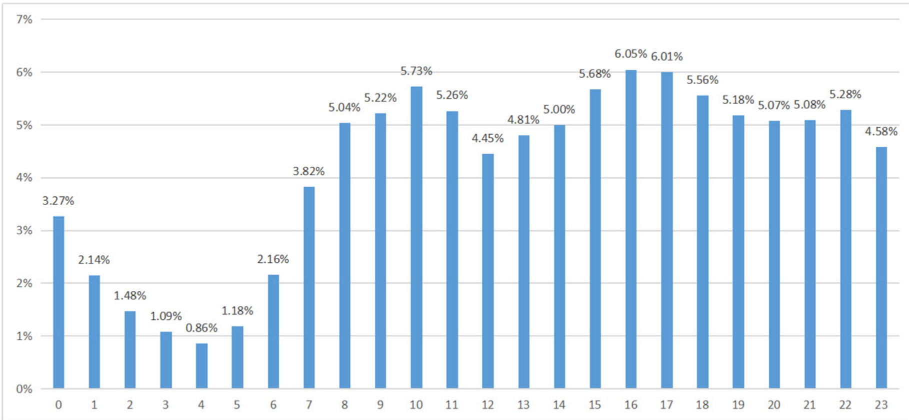

Figure 7. Total number of microblog posts released per hour

3.4.3. Comparison

In general, microblog posts on Weibo were much more positive than tweets on Twitter. People were inclined to deliver more negative information after the landfall. The difference is that the negative sentiment on Twitter was showing an expanding trend while the Weibo microblog posts started to become positive 4 days after the typhoon landed.

Additionally, public sentiment on Weibo was easily driven by the media with short-lived extreme changes. The heat peaked the day after the landfall in the US, while in China, it was the day of landfall that the heat peaked.

3.5. Number of likes

3.5.1. Hurricane Ida

Negative tweets received slightly more likes than positive ones on average, at 14.57 and 14.51 respectively.

Figure 7 shows that among negative tweets, tweets indicating fear had the least average number of likes (7.54), followed by sad tweets (12.05), and the most liked were angry tweets (18.76). Tweets showing joy were liked least in positive tweets, at 15.52, followed by surprise (20.36) and love (21.52).

As shown in Figure 8, the variance of negative tweets' likes is significantly larger than that of positive ones, with a variance of 295917 for the former and 84592 for the latter. It is mainly due to the large variance of angry tweets' likes, reaching 460384, over 4 times more than the second largest variance (108287).

ISSN: 2637-6067
DOI: 10.6918/IJOSSER.202203_5(3).0059

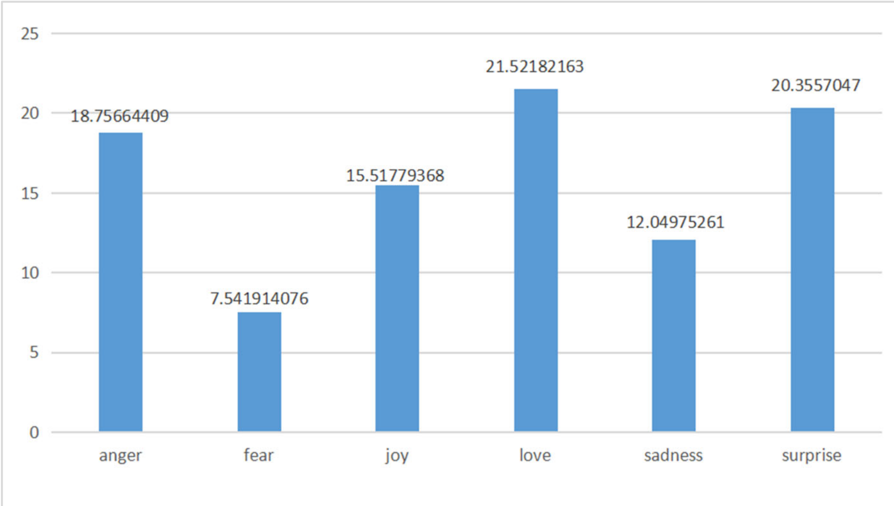

Figure 8. Average likes received by tweets about Hurricane Ida

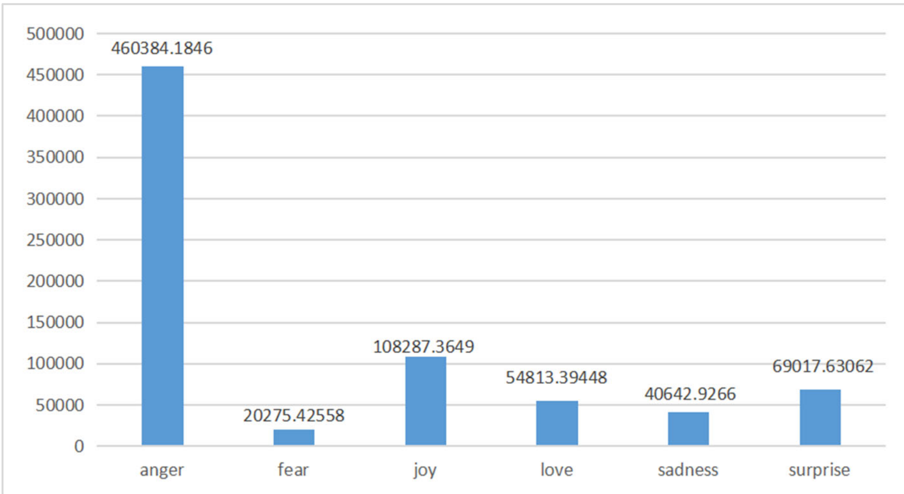

Figure 9. The variance of likes received by tweets about Hurricane Ida

3.5.2. Typhoon In-fa

Neutral, positive and negative microblog posts received 76.94, 52.24, 32.44 likes on average, with the variance of 5 352 490, 5 985 024 and 498 660 respectively.

3.5.3. Comparison

Generally speaking, positive posts were more welcomed on Weibo than Twitter. Moreover, positive posts on Weibo as well as negative posts on Twitter saw a higher variance in number of likes.

3.6. Sending sources

Figure 9 shows that iPhone and HUAWEI are the two largest post sources, significantly larger than other cell phone brands or platforms.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0059

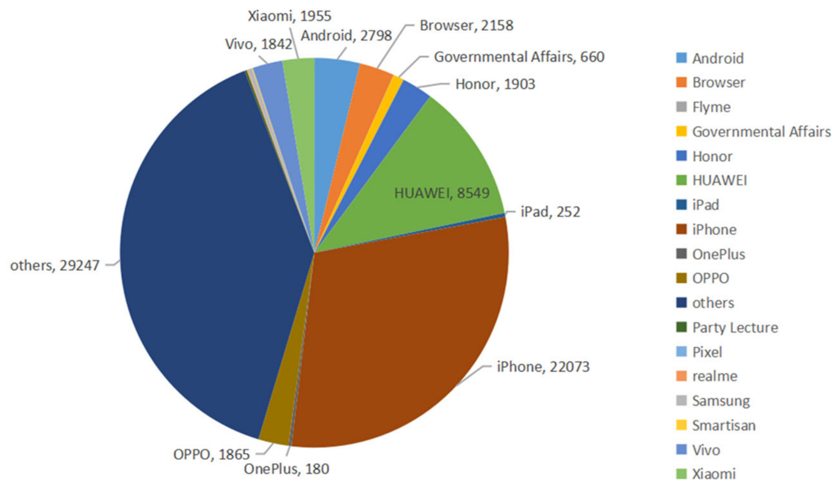

Figure 10. Sending sources of microblog posts about Typhoon In-fa

Figure 10 and 11 combined reveal that the vast majority of cell phone brands witnessed an expanded share of negative sentiment, with the largest increase being iPhone (3%). In addition to that, most other brands and platforms contain positive sentiment so that they account for 7% more positive microblog posts than negative ones.

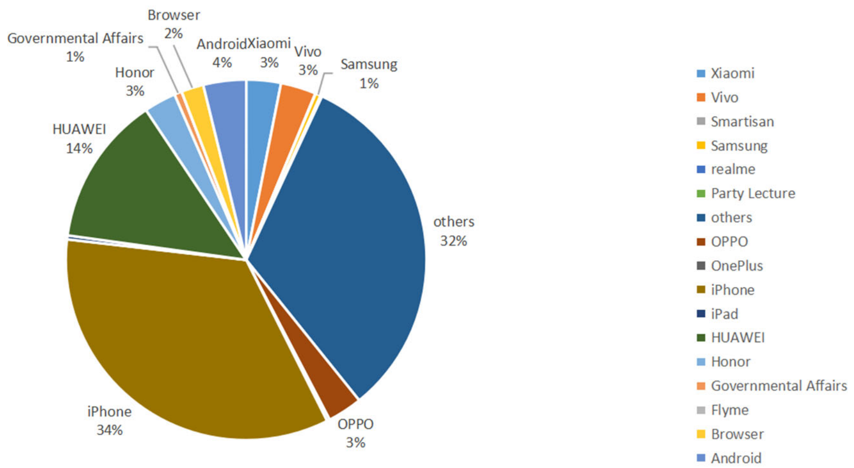

Figure 11. Sending sources of positive microblog posts about Typhoon In-fa

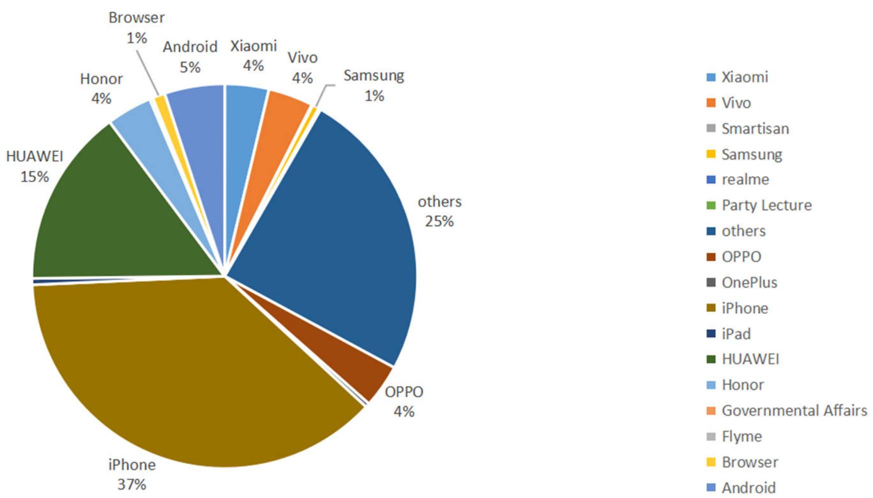

Figure 12. Sending sources of negative microblog posts about Typhoon In-fa

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0059

4. Discussion

It is abnormal to see a prominent surge of negative sentiment related to Typhoon In-fa on the evening of July 25. This may be caused by the coverage of the typhoon's influence to Henan province, where a rare flood partly caused by the very typhoon ten days ago had claimed 302 lives

However, there are also limitations to the comparison between Hurricane Ida and Typhoon In-fa. On one hand, the research used "Typhoon In-fa" as the keyword when scraping on Weibo, which filtered out some results and ended up less results compared to those collected on Twitter. This is a compromise for the limitation set by Weibo, which only allows access to a maximum of 2,000 pieces of data at any one hour. Therefore, posts sent by official media were more likely to be included than those sent by individuals, in that "Typhoon In-fa" is a formal expression. On the other hand, most posts sent by official media function as reassurances, so that positive microblog posts have become the main body since July 29. This also explains why the sending source "others", which is comprised of most official media, takes up 7% more in positive posts than negative ones.

5. Conclusion

In this study, we have found some similarities and differences in public sentiment about similar disasters in the most popular social media platforms in China and the US.

First, locations were frequently mentioned before the landfall in both countries. After the landfall, people in the US were wary, empathetic and sad while people in China cared more about the current situation and sent warning to areas in the path of the typhoon. Second, provinces where Typhoon In-fa first arrived posted the most and the population determines the number of microblog posts in the remaining areas. Third, people in both countries were generally optimistic about the imminent cyclone until the landfall. The difference is that negative sentiment has occupied Twitter since the landfall while positive one took ground again four days after the landfall on Weibo. Forth, significantly more likes were received on average by positive microblog posts on Weibo than Twitter. This is partly attributed to official media having more followers and influence than individuals. Lastly, iPhone and HUAWEI combined account for nearly half of the sending source. Microblog posts sent by official media were more likely to be positive than individuals'.

Comparisons of the geography and sending sources of this topic can be carried out if location information and sending sources are available and easy to handle on Twitter. Also, more accurate data and convincing conclusions will be collected and drawn if Weibo allows access to all history posts.

References

[1] Neppalli, V. K., et al. (2017). "Sentiment analysis during Hurricane Sandy in emergency response." International journal of disaster risk reduction 21: 213-222.

[2] Zhang, Y., et al. (2020). "Sentiment Analysis of Sina Weibo Users Under the Impact of Super Typhoon Lekima Using Natural Language Processing Tools: A Multi-Tags Case Study." Procedia Computer Science 174: 478-490.

[3] WANG Zhi-ying, L. Y.-j. (2017). "Propagation law and coping strategies for public opinions in emergency with the consideration of the government intervention." Journal of Management Sciences in China 20: 43-52+62.

[4] JustAnotherArchivist (2018), snscrape. https://github.com/JustAnotherArchivist/snscrape.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202203_5(3).0059

[5] Medhat, W., et al. (2014). "Sentiment analysis algorithms and applications: A survey." Ain Shams engineering journal 5(4): 1093-1113.

[6] Devlin, J., et al. (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.

[7] Sanh, V., et al. (2019). "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter." arXiv preprint arXiv:1910.01108.