A New Internet Public Opinion Evaluation Model: A Case Study of Public Opinions on COVID-19 in Taiwan

Sheng-Tsung Tu, Ming Chuan University, Taiwan

Louis Y. Y. Lu, Yuan Ze University, Taiwan

Chih-Hung Hsieh, Yuan Ze University, Taiwan

Chia-Yu Wu, Yuan Ze University, Taiwan

ABSTRACT

This research retrieved public opinions on the novel coronavirus pandemic with the aid of the DiVoMiner. The data were collected by setting keywords via qualitative comparative analysis (QCA) and automated computational approach, and the collected data were analyzed subsequently. The present study divided keyword collections into three categories, namely the name of diseases, government policies, and COVID-19 events. It was found the retrieved internet public opinions on COVID-19 were the largest in number and contained the least noise when the three categories of keywords appeared at the same time. Therefore, the data of internet public opinions = the name of diseases $\times$ (government policies + COVID-19 events). This research found that an event that happens daily will affect the number of internet public opinions on social media and forums after it has been reported. The strong negative emotion conveyed through the internet public opinion may turn into a positive one if the event is dealt with properly after positive focus words represent the same proportion as negative ones.

KEYWORDS

Automated Computational Approach, COVID-19, Divominer, Internet Public Opinion, Qualitative Comparative Analysis

1. INTRODUCTION

In December 2019, cluster infection of unidentified COVID-19 occurred. A novel coronavirus, which can spread across species and through human-to-human contact, was isolated and sequenced, which was named “severe acute respiratory syndrome coronavirus 2 (i.e. SARS-CoV-2). On February 11, 2020, the World Health Organization (WHO) officially named the severe infectious pneumonia caused by this virus Coronavirus Disease-19 (i.e. COVID-19). Due to its rapid spread and inevitable high infectivity, it has been classified as a major infectious disease by WHO. It has spread to 185 countries and regions (Di Gennaro et al., 2020). As of November 10, 2020, there were a total of 50,913,451

DOI: 10.4018/IJBDAH.287603

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

confirmed cases worldwide, resulting in 1,263,089 deaths according to the COVID-19 dashboard of Johns Hopkins University.

In the modern society that features information explosion, people are becoming increasingly dependent on the Internet. Against this background, the young are more willing to voice their opinions on the Internet, commenting on news, responding to the posts on social media, and participating in discussion on a forum. Thus, their remarks about an event can be regarded as Internet public opinions. Internet users' increasing adherence to the Internet enables the authorities to promote their policies via social media and traditional news media to regard social media like Facebook and Youtube as a platform to disseminate information.

The Taiwanese authorities took the lead in COVID-19 prevention because of an article posted on PTT Gossiping. On December 31st, 2019, the Internet user “nomorepipe” published the article “Suspected SARS Coronavirus Cluster Infection Broke out in Wuhan?” The netizen posted a test report concerning the coronavirus by Li Wenliang, who is claimed Coronavirus Whistleblower doctor. The article led to heated discussion, which drew the attention of the Centers for Disease Control and Prevention (CDC). Therefore, Taiwan initiated its epidemic prevention in January 1, 2020, which was earlier than other countries and regions across the world.

In brief, the present study aimed to use DiVoMiner to retrieve the data about Internet public opinions on COVID-19 between December 31, 2019, the time when public opinions on COVID-19 first appeared, and June 7, 2020, the time when Central Epidemic Command Center (CECC) “unsealed” Taiwan. Afterward, QCA and ACA were employed to set the keywords that were used to retrieve and analyze data, expecting to clarify the following questions:

1. How many reports and discussions about COVID-19 are there on the Internet, including news media, social media and forums?
2. What are the words that Internet users choose to comment on COVID-19 events?
3. How do the emotions that the Internet public opinions on COVID-19 convey change?
4. Is there a correlation between the numbers of public opinions about COVID-19 on different media?

2. INTERNET PUBLIC OPINIONS

Compared with traditional media, the Internet has fewer limitations caused by carriers. Moreover, the Internet gradually changes the way of information transmission: the information was previously disseminated from “one to one” and “one to many”, but now it is transmitted from “many to many”. On the other hand, the information flow on the Internet spreads fast but gets insufficient management. As for public opinions, they are complicated and may be antagonistic to each other. Worse still, readers' subjective judgment during information dissemination makes the media unable to play the role of “gatekeeper” like before. As a consequence, the authenticity of information on the Internet cannot be guaranteed. In this context, a great deal of ever-flowing information inevitably includes false and unconfirmed one. Therefore, opinion leaders may take Internet users to a wrong direction, and some even deliberately take public opinions to a direction that helps them achieve their goals (Huang Wei, Li Rui, & Meng Jialin, 2015).

Previous researches on “Internet public opinions” have defined it. Overall, Internet public opinions can be deemed as the emotions, ideas, opinions, attitudes and social influence that a netizen¹ has toward a social problem, public event, ideology, and morality with the Internet as the carrier and an event as the core. The rapid spread of Internet public opinions exerts a great impact on many aspects of social life. With the rapid development of the Internet, Internet public opinions form quickly and its social influence is getting increasingly larger. The influence of Internet media has far exceeded that of traditional media, such as newspapers, radio and television, which makes it hard for traditional media to continue its development. Therefore, traditional media has been undergoing adjustments.

As well, Internet public opinions change its way of presentation rapidly. In the initial stage, they were mainly manifested in news reviews, PTT forums, comments, and reposts. They are now mainly presented via Facebook and Instagram.

In fact, public opinions on the Internet are similar to that in real world, and the remarks and deeds of opinion makers on the Internet and in real world are both similar and different. In reality, residents can be classified into different classes, social groups or interest groups. In the Internet society, netizens participate in different online communities according to their own interests, preferences, and values. The difference between the two lies in the fact that in the Internet society, netizens have no labels that they bear in real world. No matter what role they play in real world, they can equally express their opinions on the Internet in an anonymous manner. The equality between communication subjects on the Internet also brings to an end the era of discourse monopoly between the government and the media. Due to the obstacles from society as well as cultural and ideological influence, people are often unable to express their emotion, wish, dissatisfaction and anger in the real world. The anonymity of netizens on the Internet provides them with the opportunity to express the dissatisfaction that they have experienced in the real world. This way, Internet public opinions come into being. Moreover, some Internet public opinions will affect the real society, causing the reaction of social subjects, such as Mass Protest over Corporal Chung-Chiu Hung, Jasmine Revolution, the Arab Spring events, etc.

The literature review reveals multiple problems as follows:

1. **Data from Limited Number of Websites:** Some of the previous researches on big data often wrote web crawlers through word patterns and programs to retrieve data from specific websites. This way, relevant data only come from given websites, which may cause research inaccuracy and loss of reliability and validity unless a research aims to measure the public opinion of given websites or Internet forums.
2. **Problems about Setting Keywords:** Previous researches in this field often used a single keyword or keyword combination (usually four to five keywords) to retrieve data about public opinions. However, if researchers set keywords this way, they tend to obtain noise data, which necessitates time-consuming data collation. Research on Internet public opinions values efficiency, while data collation may lead to the loss of timeliness that public opinion evaluation requires.
3. **Excessive Intervention:** At the initial stage of the evaluation of Internet public opinions, trained coders conducted data coding and emotional judgment in some researches. Compared with computer-based calculation, manual judgment unavoidably causes judgmental inaccuracies, thus resulting in judgmental errors. In this respect, the computer-based calculation has greater chance to improve the overall accuracy of Internet public opinion evaluation.

3. METHODOLOGY

3.1 Computer-Based Calculation

Natural language processing (NLP) of machine learning enables computers to automatically analyze massive data, including trend analysis, emotional analysis, and breaking-down-sentence analysis. Therefore, this research employed DiVoMiner to structure the data of different sources through rigorous analysis and a reliability monitoring mechanism. This way, all the data were gathered in a single platform for internal auditing and filtration, after which rigorous content analysis method, which involves real-time coding, examination, monitoring and presentation, was utilized to visualize the data and conduct valuable semantic analysis at the same time. This way, this research obtained a report that has insights and facilitates decision-making. At present, DiVoMiner collects data from major news platforms, forums, PTT, Facebook, Instagram, Youtube, etc.

The semantic machine learning model having been introduced into DiVoMiner, the textual mining and analysis platform can not only be used for manual coding, but also for machine learning coding,

multidimensional analysis, analysis of the correlation between multiple variables, cross analysis, regression analysis, statistical verification, and the creation of word clouds. DiVoMiner also has a complete mechanism that monitors coding performance, which enables it to control the efficiency and accuracy of sampling. In addition, DiVoMiner adopts different inferential statistical methods to verify the representativeness of the results.

3.2 Keyword Setting

In the present study, keyword combinations consisted of three categories, i.e. the name of diseases, government policies, and COVID-19 events, which were analyzed via QCA.

QCA was a method published by the American social science scholar Charles C. Ragin in 1984. It was first proposed in 1987 (Rihoux, 2003; Dixon-Woods et al., 2005; Rihoux, 2006; Schneider & Wagemann, 2006), but it was not widely used until 1997 (Ragin, Shulman, Weinberg, & Gran, 2003). QCA, a method that integrates quantitative and qualitative research, features an analytical model of set theory. QCA is suitable for analyzing small- and medium-sized samples, while it has been used to analyze large-sized samples in a small number of studies; it has been compared with quantitative analysis like logistic regression, and the results obtained are almost the same (Grendstad, 2007).

QCA creates a truth table based on the dichotomy of “0” and “1” for the topic-related factors. The appearance of the code 0 means the absence of representative factors whereas the appearance of the code 1 indicates the existence of the representative factors. Afterward, Boolean Logic was used to figure out the configurations (Ragin, 1987), which provides a relatively objective basis for the explanation of cause-effect relations (i.e. causation).

When three factors are used for QCA, there are eight ($2^3 = 8$) combinations. For instance, X is a dependent variable, a, b, and c are independent variables; ab represents the combinations that a and b appear simultaneously, while ac represents the combinations that a and c appear at the same time. Therefore, $X = ab + ac$ indicates that X is the connected set of the combinations ab plus the combinations ac. Meanwhile, it means that the dependent variable X will have the maximum, so the equation can also be represented as $X = a * (b + c)$, indicating that a is the necessary factor for X (Ragin, 1999a; Ragin, 1999b; Rihoux, 2006). Thus, it means that the appearance of the factor a inevitably will result in the emergence of the dependent variable X.

QCA is a qualitative research method, which necessitates full discussion based on relevant data when researchers select factors and conduct coding. Moreover, whether the obtained results are proper should be discussed as well. An advantage of QCA is that it provides a systematic analysis of complicated and massive qualitative data. In addition, researchers obtain consistent equation as long as the researchers use the same variable for analysis and the same coding. Therefore, QCA has the characteristics of universal applicability and extrapolation, and reaches a level that qualitative research could not achieve before (Rihoux, 2003; Rihoux, 2006). Moreover, QCA can not only verify regular combinations found in existing researches, but also find out accidental or abnormal factors.

Regarding the name of diseases, keyword setting experienced two stages. First, academic terms related to Wuhan pneumonia that appeared in relevant reports both home and abroad were measured, such as Wuhan pneumonia, novel coronavirus, nCoV, novel coronavirus 2019, 2019-nCoV, severe infectious pneumonia, novel coronavirus pneumonia, NCP, COVID-19, and novel corona pneumonia. Subsequently, the keywords used for the first measurement were modified, after which they were discussed by experts and scholars of the related fields. The keywords were finally set after data cleaning.

The keywords obtained after data cleaning were as follows:

Wuhan pneumonia or novel coronavirus or nCoV or 2019 novel coronavirus or 2019-nCoV or severe special infectious pneumonia or novel coronavirus pneumonia or NCP or COVID-19 or novel corona pneumonia or MERS or severe special infectious pneumonia.

As for the keyword combinations of government policies, they were set based on the categories and keywords listed on the COVID-19 epidemic prevention network established by the CECC. The keywords were as follows:

1. **Community-based Epidemic Prevention:** Home quarantine or delaying the start of school or epidemic prevention care leave or autonomous health management or travel history or contact history or centralized quarantine or large-scale rally or “epidemic prevention” hotel or “epidemic prevention” taxi or social distancing or body temperature measurement or wearing masks or closing business or control of the flow of people or new life under “epidemic prevention” or real-name system or real-name registration.
2. **Border policy:** Boarding quarantine or travel epidemic or Arrival Health Declaration Form or ban on entry or visa control or flight dedicated to Taiwanese businessmen or entry ban or prohibited entry or transfer prohibited or Taiwan-bound flight or opening dedicated flights or monitoring of the flow of people or negative or airport inspection or airport quarantine or Chinese nationality or conditional access to Taiwan or overseas students.
3. **Supplies:** Epidemic prevention items or restricted exports or Mask National Team or masks or Mask Real-Name System or alcohol or mask map or real-name masks or mask 2.0 or eMask the Name-based System for Mask Purchasing or opening exports.
4. **Relief or compensation policy:** Special Act for Prevention, Relief and Revitalization Measures for Severe Pneumonia with Novel Pathogens or relief or epidemic prevention compensation or relief scheme or living allowance.
5. **Control of medical institutions:** Restrictions from going abroad or group gatherings or access control or personnel control or easing control.
6. **Inspection / Research & Development (R & D) Policy:** Remdesivir or vaccine or quarantine or inspection or self-paid examination.

The keyword combinations for COVID-19 events were set based on relevant news media reports. The keyword combinations that appear in aforementioned categories will not be included in this category, such as masks, relief policies, etc.

1. **Delaying the Start of School:** (Senior high school and below and delaying the start of school) or (colleges and delaying the start of school) or (Students from Chinese mainland and delaying the arrival in Taiwan) or (Students from Hong Kong and Macau and suspending entries) or epidemic prevention care leave or soothing schooling program.
2. **Flight dedicated to Taiwanese businessmen:** Flight dedicated to Taiwanese businessmen in Wuhan or Flight dedicated to Taiwanese businessmen or evacuation flights or evacuation flights from Wuhan.
3. **Religious activities:** Mazu or Baishatun Mazu or Dajiamu or Dajia Mazu or Baishatun Gongtian Palace or Dajia Zhenlan Palace or Pilgrimage or Mazu circumnavigation activity.
4. **Hoarding:** Hoarding or rushing to purchase goods or panic buying or storing up food or replenishing or hoarding masks or hoarding food or rushing to buy toilet paper or toilet paper.
5. **Dunmu fleet infection:** Dunmu fleet or confirmed cases on a navy warship or Panshi Fleet.

The above three keyword combinations were used to make a truth table through QCA, as shown in Table 1.

In the present research, QCA was conducted. It was found that X, the data about Internet public opinions on COVID-19, can be calculated by the following equation:

$$X = \text{the name of diseases} \times (\text{government policies} + \text{COVID-19 events})$$

Table 1. A truth table for covid-19 keyword combinations

Code	1	2	3	4	5	6	7	8
Name of diseases	0	1	1	1	1	0	0	0
government policies	0	1	1	0	0	1	1	0
COVID-19 events	0	1	0	1	0	1	0	1

Source: collated by the present study

Consequently, the keywords used to retrieve data include two categories; one is the name of diseases while the other is about government policies and COVID-19 events. After the keyword combinations were confirmed, the evaluation of the present study was conducted based on the keyword combinations.

4. RESEARCH RESULTS

This study retrieved 1,696,010 articles related to COVID-19 from the DiVoMiner database of Internet public opinions. The first article, titled “Suspected SARS coronavirus cluster infection broke out in Wuhan?”, appeared on PTT gossiping on December 31st, 2019. The Internet user named nomorepipe posted a test report concerning the novel coronavirus by Li Wenliang, who is claimed Coronavirus Whistleblower doctor. The article resulted in heated discussion, which attracted the attention of Taiwan’s Centers for Disease Control and Prevention (CDC). As a result, Taiwan initiated its epidemic prevention on January 1, 2020, which was earlier than other countries and regions across the world. Despite the effort, Taiwan had the first confirmed COVID-19 case, which was overseas imported, on January 21, 2020, and the first local COVID-19 case on January 28, 2020.

Among the 1,680,899 articles related to COVID-19, 65.09% were published on news media (n = 1,094,044), 29.66% came from social media (n = 498,562), and 5.25% from forums and discussion boards (n = 88,293). Of 498,562 articles from social media, a dominant majority were published on Facebook (91.49%; n = 456,113), and merely 8.51% came from Youtube and Intargram (n = 45,449). Regarding news media, an average of 6,838 reports and discussions about COVID-19 (range 0-14,088) were published, while on social media, there were 3,116 posts, reposts, and discussions about COVID-19 (range 0-6,359). As for forums and discussion boards, an average of 552 articles and discussions about COVID-19 (range 0-1,210).

Regarding the number of Internet public opinions on COVID-19, as of January 19, the number of public opinions was less than 1,000 each day. However, Taiwan expanded its epidemic prevention area to airports on January 19 because it had 4 suspected cases on January 17, 2020. Since January 20, 2020, the number of Internet public opinions has exceeded 1,000 each day. Therefore, January 20 can be regarded as the starting point for the drastic growth of Internet public opinions on COVID-19.

On the other hand, since January 28, on which Taiwan had the first confirmed local case, the number of public opinions about COVID-19 reached the peak (i.e. 10,722) on news media on January 30. Moreover, the increase in the number of public opinions took on a regular pattern. During the weekdays, the number of Internet public opinions on news media all exceeded 10,000. On weekends, the number was all greater than 6,000. The changes in the number are presented in Figure 1.

Further analysis found that when a relatively large number of Internet public opinions is often attributed to the occurrence of an event, which results in a surge in Internet public opinions on the same day or the next few days. For instance, the highest point for the number of Internet public opinions on COVID-19 on a day appeared on March 19 (n = 21,511) because “hoarding chaos” happened on the same day.

Figure 1. Tendencies for internet public opinions on covid-19 (Compiled by the present research)

During the epidemic, “hoarding as soon as possible” went viral on the Internet, which led a massive number of people to rush to purchase daily necessities. On March 19, an Internet user named tombknight asked about the necessity of hoarding on PTT gossiping (Figure 2 and 3). In the post, tombknight mentioned that the elderly asked the family to store a-month-worth food and withdraw deposits from banks. The post did not trigger much discussion, and a majority of the comments showed that they did not believe that it was necessary to do so.

The analysis of keywords can reveal the event that netizens discuss on the Internet or the words that news media use in their reports during the COVID-19 pandemic. The words mentioned the most frequently include the name of the disease, such as Wuhan pneumonia ($n = 2,555,411$) and novel coronavirus ($n = 1,281,097$), the official organization, such as the Central Epidemic Command Center (CECC) ($n = 2,330,279$), and the name of hard-hit countries, like China ($n = 441,844$) and America ($n = 333,058$). The words on word clouds were frequently brought up during the COVID-19 pandemic,

Figure 2. Hoarding-related post on March 19 (Compiled by the present research)

Figure 3. Hoarding-related post on March 19 (Compiled by the present research)

as shown in Figure 4, and the top 20 words that often appear in the Internet public opinions on the COVID-19 pandemic are listed in Table 2.

Emotional analysis of the Internet public opinion on COVID-19 reveals the emotions that netizens' opinions convey, as shown in Figure 5. COVID-19 was called Wuhan Pneumonia in the beginning, which caused netizens' negative emotions to remain intense. With the decrease of confirmed cases in Taiwan, their negative emotions was getting less intense from the peak (84.18%, $n = 165$). On the other hand, officials and soldiers on Dunmu Fleet, a navy warship, were confirmed infected between March 18 and March 23, which caused negative emotions to intensify (greater than 50%). Subsequently, the pandemic was taken under control, and on May 3, no confirmed cases was found for the first time since March 23. On May 4, the lines of the proportions for the words used to convey

Figure 4. Word cloud for the keywords in internet public opinions on Covid-19 (Compiled by the present study)

Table 2. The top 20 keywords in internet public opinions on COVID-19

Keywords about Internet Public Opinions on COVID-19	Count (n)
Wuhan pneumonia	2,555,411
CECC	2,330,279
Novel coronavirus	1,281,097
Epidemic prevention	667,416
Definitive diagnosis	509,752
Government	470,083
The whole globe	458,540
The public	442,113
China	441,844
Masks	435,606
Influence	401,607
Infection	392,776
Measures	371,398
America	333,058
Work	321,279
Hygiene	317,455
Cases	292,979
Quarantine	279,158
Health	276,610
International	272,476

Source: Collated by the present study

Figure 5. Emotions in internet public opinions on COVID-19 (Compiled by the present study)

positive (38.69%, $n = 3,349$) and negative (37.75%, $n = 3,268$) emotions crossed; in other words, their percentages equaled. Since that, the positive emotions remained stronger than the negative one until June 7, the finishing time of the present study, so the ratio of P to N remained greater than 1.

Moreover, the emotions in the public opinions on COVID-19 were categorized based on time periods², which are made into perceptual maps³, as shown in Figure 6. It can be found that during the first period, the emotional perception fell in the third quadrant, which was small in the number of public opinions and low in the ratio of P/N, for the epidemic was not clear and COVID-19 was named Wuhan pneumonia in this phase. Between the second and fourth periods, the emotional perception mainly fell in the fourth quadrant, which is large in the number of public opinions and low in the ratio of P/N, for the epidemic was becoming worse and the events like “mask chaos” and “hoarding chaos” happened. In the fifth period, the emotional perception fell in the second quadrant, which was small in the number of public opinions and high in the ratio of P/N because the COVID-19 pandemic was taken under control in Taiwan, so the emotions were positive and the discussion about it decreased.

As shown in Table 3, further analysis of the emotions in the posts on news media, social media and discussion boards, can be found that news media were in a relatively neutral position ($P/N = 0.92$)

Figure 6. Emotional perception of the internet public opinions on COVID-19 (Compiled by the present study)

Table 3. Emotions in internet public opinions on COVID-19 on different media

Variables	Positive, n (%)	Neutral, n (%)	Negative, n (%)	P/N ratio
News ($n = 867,647$)	312,814 (36.05)	216,422 (24.94)	338,411 (39.00)	0.92
Social Media ($n = 476,766$)	143,085 (30.01)	71,126 (14.92)	262,555 (55.07)	0.54
Forum ($n = 79,886$)	16,213 (20.30)	24,510 (30.68)	39,163 (49.02)	0.41
Total ($n = 1,424,299$)	472,112 (33.15)	312,058 (21.91)	640,129 (44.94)	0.74

Source: Compiled by this study

when reporting COVID-19 events. As for social media, although news media repost their reports on their social media account, Internet users conveyed stronger negative emotions on social media, for they are able to freely comment on the post. With regard to the discussion board, netizens often send radical posts or comments, so its P/N was 0.41, indicating that stronger negative emotions were seen on the discussion board.

Furthermore, the analysis of the replies to some comments found that suspected netizens from Chinese mainland participated in the reply. One reply read, “Our Taiwan of China should have faith in the authorities. Everything will be fine.”, as shown in Figure 7 and 8, while the other read, “Thanks, Taiwan Province of China”, as shown in Figure 10 and 11. When these two posts appeared, Taiwanese netizens flocked to comment on the posts.

The identities of the two suspected Chinese netizens were analyzed, and it was found that they might be part of the Chinese cyber army, as shown in Figure 9 and 12. The Twitter accounts of a cyber army have two characteristics: one is that the account has few friends, while the other is that their posts are sent during office hours (Hsin-Chang Jian, & You-Ju Li, 2019). The profile of the above two accounts and the time when their posts were sent conform to the above two characteristics. That is why the present study presumes that they are part of the Chinese cyber army.

The present study conducted correlation analysis, expecting to explore the correlation between different media, as shown in Table 4. The Pearson correlation analysis revealed that the correlation coefficients fell between 0.842 and 0.954, indicating highly positive correlations between the media. In addition, it can be found that when the number of Internet public opinions on COVID-19 increased on a media, the number on the other two showed a similar tendency.

Figure 7. The posts sent by a suspected account of the Chinese cyber army (Compiled by the present research)

Figure 8. The posts sent by a suspected account of the Chinese cyber army (Compiled by the present study)

Figure 9. An account suspected to be part of the Chinese cyber army (Compiled by the present research)

Figure 10. The posts from a suspected account of the Chinese cyber army (Compiled by the present research)

Figure 11. Posts sent by a suspected account of the Chinese cyber army (Compiled by the Present Study)

5. CONCLUSION

The analysis of Internet public opinions on COVID-19 reveals that the public opinions mainly appear on news media, which is followed by social media like Facebook, Youtube, and Instagram, and forums respectively. During the research period, there were an average of 6,838 reports and discussions about COVID-19 on news media. Since January 30, 2020, on which the first peak of the number of

Figure 12. A suspected account of the Chinese cyber army (Compiled by the present study)

Table 4. Correlation between internet media regarding tendency of public opinions

Media	News	Social Media	Forums
News	-	0.954**	0.842**
Social Media	0.954**	-	0.941**
Forums	0.845**	0.941**	-

** P<0.01

public opinions appeared, the Internet public opinions on COVID-19 presented a regular pattern. On weekdays, the number of public opinions on news media exceeded 10,000, while the number exceeded 6,000 on weekends, for related organizations conducted press conferences and made announcements when the COVID-19 pandemic began, which influenced the reports of news media.

The analysis of the keywords related to the COVID-19 pandemic revealed that the focus words were highly related to COVID-19, either in the reports of news media or on the discussion board, such as Wuhan pneumonia, CECC, novel coronavirus, epidemic prevention, and definitive diagnosis.

Further analysis of the emotion conveyed by keywords revealed that some words had been deemed as ones that convey negative emotions like Wuhan pneumonia, but they appeared in COVID-19 posts that conveyed both positive and negative emotions. Therefore, it can be regarded as a word that convey neutral emotions.

In terms of the emotions of public opinions on COVID-19, focus words used to convey negative emotions accounted for 84.18% because COVID-19 is a worldwide pandemic and the virus was first believed to originate from China. The lines of the percentages for the focus words that convey positive and negative emotions crossed on May 4, 2020, because Taiwan gradually took under control the

COVID-19 pandemic, and the confirmed cases only occurred in a small scale, either for overseas exported cases, local cases or for cluster infections on Dunmu Fleet. Subsequently, the percentage of the focus words that convey positive emotions remained higher than that of negative ones between May 5, 2020 and June 7, 2020, the time when CECC announced “unsealing Taiwan”, and the ratio of P to N remained greater than 1.

Lastly, the peaks of the number of Internet public opinions on COVID-19 were compared with relevant events, showing that the events, such as the occurrence of confirmed cases, rushing to purchase masks, and hoarding chaos, all increased the discussion on COVID-19 online. The correlation analysis revealed that news media, social media and forums were highly positive correlated. Therefore, it can be concluded that COVID-19 events affect the changes in the number of Internet public opinions.

REFERENCES

- Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A. S., & Buyya, R. (2014). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, 79–80, 3–15.
- Di Gennaro, F., Pizzol, D., Marotta, C., Antunes, M., Racalbutto, V., Veronese, N., & Smith, L. (2020). Coronavirus Diseases (COVID-19) Current Status and Future Perspectives: A Narrative Review. *International Journal of Environmental Research and Public Health*, 17(8), 2690. doi:10.3390/ijerph17082690 PMID:32295188
- Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising Qualitative and Quantitative Evidence: A Review of Possible Methods. *Journal of Health Services Research & Policy*, 10(1), 45–53. doi:10.1177/135581960501000110 PMID:15667704
- Du, S. (2020). *The Tendency and Rise and Fall of Taiwanese Consciousness of Unification and Independence: A Comparison of Three Search Engines of the Databases*. Hanlu Book Publishing.
- Grendstad, G. (2007). Causal Complexity and Party Preference. *European Journal of Political Research*, 46(1), 121–149. doi:10.1111/j.1475-6765.2006.00689.x
- Huang, W., Li, R., & Meng, J. (2015). Study on Dissemination Elements and Operational Mechanism of Multimedia Network Public Opinion under the Big Data Environment. *Library and Information Service*, 59(21), 38–44.
- Jian, X., & Li, Y. (2019). *How Twitter Recognizes Cyber Army Accounts?* READr: https://www.readr.tw/post/2029?fbclid=IwAR1lnz-5V_Esl8iRdKPZOLTjN8hGWyLEGt3wE7QLSfLQj2HZH-GSSDc9m50
- Johns Hopkins University. (2020). *COVID-19 Dashboard*. <https://coronavirus.jhu.edu/map.html>
- Ministry of Health and Welfare. (2020). *COVID-19 Epidemic Prevention Network*. <https://covid19.mohw.gov.tw/ch/mp-205.html>
- Nan, F., Suo, Y., Jia, X., Wu, Y., & Shan, S. (2018). Real-Time Monitoring of Smart Campus and Construction of Weibo Public Opinion Platform. *IEEE Access: Practical Innovations, Open Solutions*, 6, 76502–76515. doi:10.1109/ACCESS.2018.2883799
- Ragin, C. C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press.
- Ragin, C. C. (1999a). The distinctiveness of case-oriented research. *Health Services Research*, 34(5), 1137–1151. PMID:10591277
- Ragin, C. C. (1999b). Using qualitative comparative analysis to study causal complexity. *Health Services Research*, 34(5), 1225–1239. PMID:10591281
- Ragin, C. C., Shulman, D., Weinberg, A., & Gran, B. (2003). Complexity, Generality, and Qualitative Comparative Analysis. *Field Methods*, 15(4), 323–340. doi:10.1177/1525822X03257689
- Rihoux, B. (2003). Bridging the Gap between the Qualitative and Quantitative Worlds? A Retrospective and Prospective View on Qualitative Comparative Analysis. *Field Methods*, 15(4), 351–365. doi:10.1177/1525822X03257690
- Rihoux, B. (2006). Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods: Recent Advances and Remaining Challenges for Social Science Research. *International Sociology*, 21(5), 679–706. doi:10.1177/0268580906067836
- Schneider, C. Q., & Wagemann, C. (2006). Reducing complexity in Qualitative Comparative Analysis (QCA): Remote and Proximate Factors and the Consolidation of Democracy. *European Journal of Political Research*, 45(5), 751–786. doi:10.1111/j.1475-6765.2006.00635.x
- Sourcing Big Data. (2019). uMiner Manual. Taipei: Sourcing. *Big Data*.
- Zhang, Y. (2009). The Definition and Differentiation of “Public Opinion” and Relevant Concepts. *Zhejiang Academic Journal*, 2009(3), 182–184.

ENDNOTES

- ¹ Netizen refers to the person who initiates and participates in the activities that form public opinions. It was proposed by Michael Hauben, who believes that netizens are a group of Internet users who have a community consciousness and behavioral connections with each other. “Community” here is not defined by the general sense of geographical areas.
- ² This research divided the periods of Internet public opinions based on times. The first period fell between December 31 and January 31, the second fell in February, the third fell in March, the fourth fell in April, and the fifth fell between May 1 and June 7.
- ³ The X-axis of the perceptual map presents the tendency of the public opinion while the Y-axis stands for the ratio of P to N. The average number of public opinions ($n = 10,600$) and the neutral emotion P/N equals to 1, which was set as the origin of the coordinates, and accordingly, other related values were used to set coordinates and made into perceptual maps.