Journal of Engineering Sciences
Vol 13, Issue 01, Jan /2022
ISSN NO:0377-9254

CHARACTERIZING THE PROPAGATION OF SITUATIONAL INFORMATION IN SOCIAL MEDIA DURING COVID-19 EPIDEMIC: A CASE STUDY ON WEIBO

[1]Gali Bhavya Siva Naga Rajeswari, [2]Mr.V. Rajesh Babu
[1]M.Tech [CSE] , Eluru college of Engineering and Technology, Duggirala - 534004, Andhra Pradesh.
[2]Professor, Eluru college of Engineering and Technology, Duggirala - 534004, Andhra Pradesh.

Abstract— During the ongoing outbreak of coronavirus disease (COVID-19), people use social media to acquire and exchange various types of information at a historic and unprecedented scale. Only the situational information are valuable for the public and authorities to response to the epidemic. Therefore, it is important to identify such situational information and to understand how it is being propagated on social media, so that appropriate information publishing strategies can be informed for the COVID-19 epidemic. This article sought to fill this gap by harnessing Weibo data and natural language processing techniques to classify the COVID-19-related information into seven types of situational information. We found specific features in predicting the reposted amount of each type of information. The results provide data-driven insights into the information need and public attention.

Index Terms— COVID-19, crisis information sharing, infectious disease, information propagation, social media, social net-work analysis.

I. INTRODUCTION

BURST out in Wuhan, China, the ongoing outbreak of coronavirus disease (COVID-19) has caused regional and global public health crisis [1]. During a crisis like the COVID-19 epidemic, the public tends to social media platforms to acquire needed information and exchange their opinions [2], [3]. There are many different types of information on social media platforms, and the situational information, the information that helps the concerned authorities or individuals to understand the situation during emergencies (including the actionable information such as help seeking, the number of affected people) [4], is useful for the public and authorities to guide their responses [5], [6]. Identifying these types of information and predicting its propagation scale would benefit the concerned authorities to sense the mood of the public, the information gaps between the authority and the public, and the information need of the public. It would then help the authorities come up with proper emergency response strategies [6]. The existing studies have not yet agreed on the definition of situational information. Some categorized help seeking and donations as situational information but neglecting other types of situational information such as criticism from the public, which reveals the concerns of the public, and emotional support, which reveals the empathy of others to the victims [7]. However, it is necessary to identify situational information in a comprehensive manner. For example, identifying critism related information can help the authorities learn the main concerns of the public and come up with proper responses. Identifying emotional support information could help the authorities learn the social support patterns of social media users and take better advantage of such voluntary resources. Moreover, it is critical to identify the key features that can predict the propagation scale of situational information to ensure that the authorities can publish different types of situational information based on the needs of the public. Filling such information need is critical during sudden epidemics such as COVID-19. To fill these research gaps, we used COVID-19-related discussions on Sina Weibo, the major microblogging site in China (the Chinese equivalent of Twitter) to answer the following research questions:

RQ1. How to identify and categorize the situational information in social media?

RQ2. What is the various predictability of features of the propagation scale of different types of situational information?

II. LITERATURE SURVEY

Social media platforms are widely used by people to share information in different situations [8]–[12]. During a crisis, rich situational information is generated by social media users [4], [7], [13]. However, different researchers categorized situational information into different types [7]. For example, Rudra et al. [7] defined situational information as those notifications of the casualties or injured/stranded people or helping relief operations and categorized sympathizing with the victims, praising or criticizing the relief operation, post analysis of the reasons the crisis happens, and donation-related informaIon into non situational information. While in the study by Vieweg (2012) [14], the non situational information defined by Rudra et al. [7] was categorized into situational information. Specifically, she categorized situational information into social environment information, built environment information, and physical environment information. The social environment information contains advice, caution, evacuation, fatality, injury, medical attention, people missing, and offering help. Built environment information contains damages caused by the crisis and the status of infrastructures. Physical environment information includes environmental impact, general area information (status of the hazard), and general hazard information (e.g., weather report) [14]. Moreover, Imran et al. [15] further categorized the situational information into seven types such as caution and advice, casualties and damage, donations of money, goods, or services, people missing, found, or seen, and information source based on Vieweg's work. Mukkamala and Beck [4] categorized 10 types of situational information including initial information about the disaster, situation updates, criticism about insufficient attention, moral support, preparations, criticism and control rumors, help request, offering help, self-organizing support, and active volunteerism. According to Vieweg [14], situational information is the posts which provide "tactical, actionable information that can aid people in making decisions, advise others on how to obtain specific information from various sources, or offer immediate post-impact help to those affected by the mass emergency." Sharing such information helps the concerned authorities and individuals understand the crisis and guide their behaviors [7]. Based on this definition, this study categorized seven types of COVID-19-related information as situational information:

1) caution and advice;

2) notifications and measures been taken;

3) donations of money, goods, or services;

4) emotional support;

5) help seeking;

6) doubt casting and criticizing; and

7) counter-rumor. Fig. 1 presents the content types. Specifically, caution and advice information notifies the public to help them protect themselves from the harm of the virus. Notifications (situation updates or casualties and damages) of COVID-19 tell the public about the details of the epidemic, and sharing this type of information helps them learn the situation and

Vol 13, Issue 01, Jan /2022
ISSN NO:0377-9254

helps others ease the anxiety of insufficient information [13], [15]. Donation (offering help) information helps the public especially those who need help

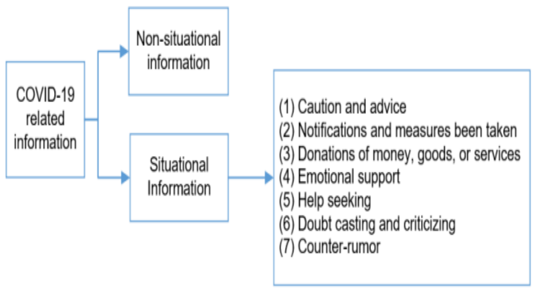

Fig. 1. Types of COVID-19-related information on Sina Weibo.

to learn what kinds of help are available [15]. Emotional support information shows positive effect on the victims to recover from the emotional harm as a result of the epidemic, and sharing this type of information helps others obtain the collective support and feel the empathy [4]. Help-seeking- related posts spread information about immediate help or aid during the epidemic, and sharing this type of information helps authorities and individuals obtain the help or aid [17]. In addition, doubt casting and criticizing information often discusses sociopolitical causes and implications of and responsibility for crisis, and sharing such information helps others vault the validity of information or enhance their understanding of the epidemic [16], [23]. Counter-rumor information also helps the public learn the truth and lessen the confusion caused by rumors [4]. Given that sharing these seven types of information benefits the efficiency of disaster/crisis relief processes, we categorized them into situational information.

III. METHODOLOGY

Several features were applied to predict the propagation scale of social media content in crisis including the content features and user features. Content features mainly contain the following: whether URL/Hashtag is contained, the publishing timing of the content, and the content's length have been recognized to positively affect the reposted amount of social media content [18]–[20]. Moreover, the content type may also significantly affect the reposted amount of information for that previous study found that users show distinct information needs and interactions when sharing different types of information [21]. Besides, emotions of the content also affect the information propagation scale of information [20]. For example, Berger and Milkman [19] characterized the effects of emotionality, positivity, awe, anger, anxiety, and sadness words in the content on the virality of social media content. From the perspective of social media users, whether users have a higher number of followers or followees (proxies of the social capital) and whether they are verified users or not also positively affect the propagation scale of the post in social media [26], [27]. Another user-related feature is users' location, and if users are from the developed cities, they are more likely to attract more social capital which will enlarge their influence upon others [28]. In addition, in the case of disasters, people are more likely to share the posts from users who are eyewitnesses of the event and located near the event [29]. Besides, affiliations and perceptions of people could also affect users' information sharing behavior [22]–[25]. As suggested

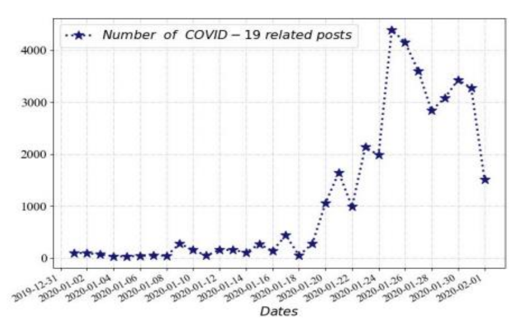

Fig. 2. Number of COVID-19-related posts each day.

Journal of Engineering Sciences

Vol 13, Issue 01, Jan /2022

ISSN NO:0377-9254

IV. METHODOLOGY

4.1 DATASET

This article collected Weibo posts using the keywords “New coronavirus ( )” and “unknown pneumonia ( )” using the Weibo API after the start of the COVID-19 epidemic in Wuhan. According to a report of the authorities on February 17, 2020, this virus has taken 1772 persons’ lives and with more than 70 000 people infected. We collected the following attributes of the Weibo data from December 30, 2019, to February 01, 2020: 1) the total number of posts is 36 746; 2) posts’ attributes including the creating time of the post, the distinct ID of the post, text content, and the length of the post, reposted/comments/likes, and the amount of the post; and 3) user attributes such as user id, tagged location, verified, follower counts, and followee counts. In addition, we plotted the number of COVID-19-related posts within the time span by date to reveal the evolution trend of the public opinion as shown in Fig. 2. The results show that public opinion began to spread with increasing speed on January 19, peaked on January 25, and started declining since February 1. This suggests that the authorities need to pay more attention to the content types in early stage of the epidemic to understand the needs of the public and adjust their information publishing strategies properly

4.2 Situational Information Classification

Various natural language processing methods were used to classify social media content into several types [30], [31]. Specifically, we use supervised learning methods such as support vector machines (SVM), naive Bayes (NB), and random forest (RF) to learn the types of unlabeled data based on labeled data [21], [30], [31]. The classification process is as follows

TABLE I

MANUALLY LABELED RESULTS OF THE SAMPLE DATA

Types	Names and definitions	Manual label#
Type 1	Caution and advice: precautions to explain the face of the epidemic should pay attention to what aspects, such as frequent hand washing, wearing masks, less out of the door or the recommendations of responding to the outbreak of the crisis.	526
Type 2	Notifications or measures been taken: outbreak announcements or the measurement already taken by the relevant departments, such as how many cases have occurred, the characteristics of the virus, material reserves, etc.	957
Type 3	Donation of money, goods, or services: donations or wishes to donate materials, money, or services for outbreak prevention and control.	182
Type 4	Providing emotional support: praise or show sympathy to others such as medical team, public welfare organizations, celebrities, and the masses who supporting Wuhan.	255
Type 5	Help seeking: (a) Medical institutions, individuals, etc. to seek support needs, etc. (b) Seek emotional support such as to seek comfort, or to express depression, etc.	262
Type 6	Doubt casting and criticizing: to question local government officials for inaction, the central government, the Red Cross and other related initiatives, or some of the public to mislead others.	522
Type 7	Refute rumors: in response to recent rumors.	52
Type 8	Non-situational information: information that are not related to the crisis.	244

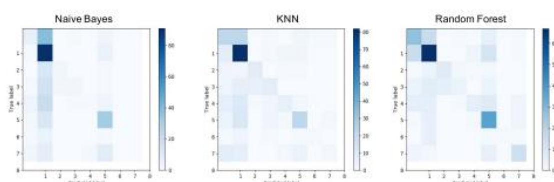

Fig. 3. Confusion matrixes of different classifiers.

First, we randomly sampled 3000 COVID-19-related posts from the collected data sets. Second, three graduate students manually labeled the type of information of the 3000 sampled data. Third, we calculated the Cohen’s Kappa value of the three coders and the value was 0.81 Cohen’s Kappa value, which is larger than 0.8 and indicates the convincible of labeling results. Table I presents the labeling results and the definitions of each type of information. Fourth, the SVM, NB, and RF classifiers were trained using randomly sampled

Vol 13, Issue 01, Jan /2022
ISSN NO:0377-9254

85% (2550) data, and using the rest of the 450 to test the accuracy of each classifier, the average accuracy of these classifiers was 0.54, 0.45, and 0.65 respectively. Since we have eight types, we further plotted the confusion matrixes of each classifier as shown in Fig. 3. The results indicate that RF classifier performs the best. Fifth, we choose the RF classifier to automatically label the rest of the data using all the 3000-labeled data as training set. Sixth, we summarized the five attributes as shown in Table II to reveal the information needs of the public in general. The attributes including the total (average) number of posts, the verified users' amount (the proportion of the verified

TABLE II

SUMMARY OF THE ATTRIBUTES OF EACH TYPE OF INFORMATION

Features	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7
affect	5.324	4.417	3.870	6.269	7.021	6.791	3.802
posemo	2.536	1.964	2.323	3.867	2.769	2.494	1.308
negemo	2.369	2.051	0.831	1.826	1.369	3.470	2.152
anx	0.450	0.454	0.236	0.373	0.339	0.677	0.218
anger	0.232	0.215	0.115	0.303	0.150	0.703	0.298
sad	0.159	0.148	0.115	0.208	0.150	0.427	0.063
percept	2.377	1.865	1.355	2.429	1.243	2.601	1.433
see	1.218	0.775	0.668	1.175	0.460	0.799	0.369
hear	0.451	0.471	0.326	0.625	0.375	0.921	0.698
feel	0.391	0.350	0.188	0.309	0.212	0.458	0.172
affiliation	1.796	1.501	2.328	2.503	6.350	1.906	2.668
reward	2.297	2.034	2.041	2.270	1.663	2.421	1.683
risk	3.410	3.113	3.171	3.785	2.563	5.304	3.254
drives	0.308	0.331	0.264	0.562	0.306	0.425	0.135
achieve	1.635	1.230	0.956	1.264	1.624	1.558	0.986
power	1.796	1.501	2.328	2.503	6.350	1.906	2.668
Verified	0.712	0.623	0.800	0.573	0.681	0.274	0.624
Followers(log)	10.783	10.004	11.742	9.510	10.298	7.367	10.153
Followees (log)	6.131	6.084	6.268	5.994	6.059	5.768	6.065
BigCity	0.145	0.133	0.172	0.131	0.093	0.113	0.133
NearCity	0.040	0.042	0.053	0.060	0.096	0.066	0.024
Hash	0.593	0.687	0.650	0.694	0.594	0.423	0.412
URL	0.000	0.000	0.000	0.000	0.000	0.000	0.000
length	54.798	53.135	58.083	42.649	49.181	47.742	51.364
hours	655.32	650.725	692.757	651.789	689.316	664.554	626.003

users), the total (average) number of followers and followees of the users, and the total (average) reposted amount of the posts in each type of information. Table II shows that during the COVID-19 epidemic, there are more verified users involved in the spreading process of Type 2 information (notifications and measures been taken). The verified users dominated the propagation of Type 3 information (donations of money, goods, and services). Type 5 information (help seeking) attracted the largest amount of average attitudes. Type 6 information (doubt casting and criticizing) have the second largest amount in our data set and a larger proportion of users involved in the propagation of this type of information. Such findings reveal the necessity to identify further information release strategies to this type of information as suggested by Atlani-duault et al. [31]. In addition, the Type 8 information (nonsituational) in our data set has the largest total amount of reposted amount. By tracing back to the data set, we found that the dominating post was posted by "CCTV News," which published this information on January 24, 2020 (Chinese New Year's Eve) to send greetings to all the users which attracted 2 774 936 reposted amounts (94.5% of all the reposted amount). After deleting this outlier, 1 62 497 total reposted amount and 56.9 average reposted amount are left. Then this nonsituational information becomes the least popular information compared with the situational information. It indicates that in this crisis, users are more likely to repost the situational information rather than the nonsituational information

V. CONCLUSION

The findings of this article indicate the necessity of using different information publishing strategies for different types of situational information. The selected features for different types of situational information could also help the authorities learn how to organize their COVID-19-related posts to enlarge or decrease the reposted amount of their posts. In addition, the definitions of the situational information will be useful for researchers or practitioners who aim to build effective social-media-based emergence response programs and crisis information systems. This article has limitations. First, we only obtained a subset of social media data for the constrains of Sina API. In future we will collaborate with the data providers to obtain a larger amount of data. Second, in training the classifiers to identify the content types of situational information, we only trained three traditional NLP-based classifiers due to limited data. In future, we will use more data and train deep learning methods to identify the content types with higher accuracy [34], [35]. Third, the manual labeling of the crisis data is time-consuming and might influence the efficiency of characterizing crisis information sharing. In future, we plan to apply automatic labeling methods to avoid this limitation [6], [36].

REFERENCES

[1] J. T. Wu, K. Leung, and G. M. Leung, “Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study,” Lancet, vol. 395, no. 10225, pp. 689–697, Feb. 2020.

[2] P. Burnap et al., “Tweeting the terror: Modelling the social media reaction to the woolwich terrorist attack,” Social Netw. Anal. Mining, vol. 4, no. 1, pp. 1–14, Dec. 2014.

[3] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, “Microblogging during two natural hazards events,” in Proc. 28th Int. Conf. Hum. Factors Comput. Syst. (CHI), 2010, p. 1079. 2010

[4] A. Mukkamala and R. Beck, “The role of social media for collective behaviour development in response to natural disasters,” in Proc. 26th Eur. Conf. Inf. Syst. Beyond Digit. Facet. (ECIS), 2018, pp. 1–18.

[5] M. Martínez-Rojas, M. D. C. Pardo-Ferreira, and J. C. Rubio-Romero, “Twitter as a tool for the management and analysis of emergency situations: A systematic literature review,” Int. J. Inf. Manage., vol. 43, pp. 196–208, Dec. 2018.

[6] L. Yan and A. J. Pedraza-Martinez, “Social media for disaster management: Operational value of the social conversation,” Prod. Oper. Manage., vol. 28, no. 10, pp. 2514–2532, Oct. 2019.

[7] K. Rudra, S. Ghosh, N. Ganguly, P. Goyal, and S. Ghosh, “Extracting situational information from microblogs during disaster events: A classification-summarization approach,” in Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM), 2015, pp. 583–592.

[8] F.-Y. Wang, D. Zeng, Q. Zhang, J. A. Hendler, and J. Cao, “The Chinese–Human Flesh-Web: The first decade and beyond,” Chin. Sci. Bull., vol. 59, no. 26, pp. 3352–3361, Sep. 2014.

[9] Q. Zhang, F.-Y. Wang, D. Zeng, and T. Wang, “Understanding crowdpowered search groups: A social network perspective,” PLoS ONE, vol. 7, no. 6, 2012, Art. no. e39749.

[10] Q. Zhang, D. Zeng, F.-Y. Wang, and T. Wang, “Under-standing crowd-powered search groups: A social network perspective,” PLoS ONE. vol. 7, no. 6, Jun. 2012, Art. no. e39749, doi: 10.1371/journal.pone.0039749.

[11] F.-Y. Wang et al., “A study of the human flesh search engine: Crowdpowered expansion of online knowledge,” Computer, vol. 43, no. 8, pp. 45–53, Aug. 2010.

[12] Q. Zhang, D. Zeng, F.-Y. Wang, R. Breiger, and J. A. Hendler, “Brokers or bridges? Exploring structural holes in a crowdsourcing system,” IEEE Comput., vol. 49, no. 6, pp. 56–64, Jun. 2016, doi: 10.1109/MC.2016.166.

[13] R. Dutt, M. Basu, K. Ghosh, and S. Ghosh, "Utilizing microblogs for assisting post-disaster relief operations via matching resource needs and availabilities," Inf. Process. Manage., vol. 56, no. 5, pp. 1680–1697, Sep. 2019.

[14] S. E. Vieweg, "Situational awareness in mass emergency: A behavioral and linguistic analysis of microblogged communications," Ph.D. dissertation, ATLAS Inst., Univ. Colorado Boulder, Boulder, CO, USA, 2012, pp. 1–300.

[15] M. Imran, C. Castillo, P. Meier, and F. Diaz, "Extracting information nuggets from disaster- related messages in social media," in Proc. Iscram, May 2013, pp. 791–800.

[16] B. Takahashi, E. C. Tandoc, and C. Carmichael, "Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines," Comput. Hum. Behav., vol. 50, pp. 392–398, Sep. 2015.

[17] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network," in Proc. IEEE 2nd Int. Conf. Social Comput., Aug. 2010, pp. 177–184.

[18] O. Tsur and A. Rappoport, "What's in a hashtag? Content based prediction of the spread of ideas in microblogging communities," in Proc. 5th ACM Int. Conf. Web Search Data Mining (WSDM), 2012, pp. 643–652. [19] J. Berger and K. L. Milkman, "What makes online content go viral?" J. Mark. Res., vol. 49, no. 2, pp. 192–205, 2012.

[20] L. Li, Q. Zhang, J. Tian, and H. Wang, "Characterizing information propagation patterns in emergencies: A case study with Yiliang earthquake," Int. J. Inf. Manage., vol. 38, no. 1, pp. 34–41, Feb. 2018.

[21] R. M. Seyfarth and D. L. Cheney, "Affiliation, empathy, and the origins of theory of mind," Proc. Nat. Acad. Sci. USA, vol. 110, no. 2, pp. 10349–10356, Jun. 2013.

[22] C. W. Scherer and H. Cho, "A social network contagion theory of risk perception," Risk Anal., vol. 23, no. 2, pp. 261–267, Apr. 2003.

[23] N. Bhuvana and I. Arul Aram, "Facebook and Whatsapp as disaster management tools during the Chennai (India) floods of 2015," Int. J. Disaster Risk Reduction, vol. 39, Oct. 2019, Art. no. 101135.

[24] L. Bui, "Social media, rumors, and hurricane warning systems in Puerto Rico," in Proc. 52nd Hawaii Int. Conf. Syst. Sci., 2019, pp. 2667–2676.

[25] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Fame for sale: Efficient detection of fake Twitter followers," Decis. Support Syst., vol. 80, pp. 56–71, Dec. 2015.

[26] M. Hofer and V. Aubert, "Perceived bridging and bonding social capital on Twitter: Differentiating between followers and followees," Comput. Hum. Behav., vol. 29, no. 6, pp. 2134–2142, Nov. 2013.

[27] M. J. Stern, A. E. Adams, and S. Elsasser, "Digital inequality and place: The effects of technological diffusion on Internet proficiency and usage across rural, suburban, and urban counties," Sociol. Inquiry, vol. 79, no. 4, pp. 391–417, Nov. 2009.

[28] S. Yardi and D. Boyd, "Tweeting from the Town Square?: Measuring geographic local networks," Methods, vol. 77, no. 3, pp. 194–201, 2010.

[29] A. C. E. S. Lima, L. N. de Castro, and J. M. Corchado, "A polarity analysis framework for Twitter messages," Appl. Math. Comput., vol. 270, pp. 756–767, Nov. 2015.

Vol 13, Issue 01, Jan /2022
ISSN NO:0377-9254

[30] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?" IEEE Trans. Dependable Secure Comput., vol. 9, no. 6, pp. 811–824, Nov. 2012.

[31] L. Atlani-Duault, J. K. Ward, M. Roy, C. Morin, and A. Wilson, "Tracking online heroisation and blame in epidemics," Lancet Public Health, vol. 5, no. 3, pp. e137–e138, Mar. 2020.

[32] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," J. Lang. Social Psychol., vol. 29, no. 1, pp. 24–54, Mar. 2010.

[33] C. K. Chung and J. W. Pennebaker, "Linguistic inquiry and word count (LIWC)," Appl. Nat. Lang. Process., vol. 2015, pp. 206–229, Feb. 2012.

[34] C. Li, J. Yi, Y. Lv, and P. Duan, "A hybrid learning method for the datadriven design of linguistic dynamic systems," IEEE/CAA J. Automatica Sinica, vol. 6, no. 6, pp. 1487–1498, Nov. 2019.

[35] Y. Xia, H. Yu, and F.-Y. Wang, "Accurate and robust eye center localization via fully convolutional networks," IEEE/CAA J. Automatica Sinica, vol. 6, no. 5, pp. 1127–1138, Sep. 2019.