COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model

TIANYI WANG[1], **KE LU**[2], **KAM PUI CHOW**[1], **AND QING ZHU**[3]
[1]Department of Computer Science, The University of Hong Kong, Hong Kong
[2]Department of Social Work and Social Administration, The University of Hong Kong, Hong Kong
[3]Department of Cardiology, Qilu Hospital of Shandong University, Jinan 250012, China

Corresponding authors: Kam Pui Chow (chow@cs.hku.hk) and Qing Zhu (2691183789@qq.com)

ABSTRACT Coronavirus disease 2019 (COVID-19) poses massive challenges for the world. Public sentiment analysis during the outbreak provides insightful information in making appropriate public health responses. On Sina Weibo, a popular Chinese social media, posts with negative sentiment are valuable in analyzing public concerns. 999,978 randomly selected COVID-19 related Weibo posts from 1 January 2020 to 18 February 2020 are analyzed. Specifically, the unsupervised BERT (Bidirectional Encoder Representations from Transformers) model is adopted to classify sentiment categories (positive, neutral, and negative) and TF-IDF (term frequency-inverse document frequency) model is used to summarize the topics of posts. Trend analysis and thematic analysis are conducted to identify characteristics of negative sentiment. In general, the fine-tuned BERT conducts sentiment classification with considerable accuracy. Besides, topics extracted by TF-IDF precisely convey characteristics of posts regarding COVID-19. As a result, we observed that people concern four aspects regarding COVID-19, the virus Origin (Gamey Food, 3.08%; Bat, 2.70%; Conspiracy Theory, 1.43%), Symptom (Fever, 2.13%; Cough, 1.19%), Production Activity (Go to Work, 1.94%; Resume Work, 1.12%; School New Semester Beginning, 1.06%) and Public Health Control (Temperature Taking, 1.39%; Coronavirus Cover-up, 1.26%; City Shutdown, 1.09%). Results from Weibo posts provide constructive instructions on public health responses, that transparent information sharing and scientific guidance might help alleviate public concerns.

INDEX TERMS COVID-19 sensing, public health, sentiment classification, social media in China.

I. INTRODUCTION

Coronavirus Disease 2019 (COVID-19), caused by a new coronavirus with higher reproductivity than SARS [1], first emerged in the People's Republic of China in December 2019 [2]. Early outbreak data grew rapidly at an exponential rate [3], and human-to-human transmission also occurred [4], [5], which brought severe challenges to China and the whole world. Soon vicarious traumatization caused by COVID-19 was found spreading in members of medical teams united in aiding the COVID-19 control and general public [6]. Social media has been found as a key platform for the public on information gathering and social learning to manage uncertainty and risks during a public health crisis. Gui *et al.* [7] investigated public concerns to the Zika virus crisis and reported mechanisms of personal risk assessment and travel-related decision making during the

crisis. Meanwhile, social media have been widely used by public health professionals for epidemiological monitoring and understanding public reactions to urgent public health issues. Pei *et al.* [8] developed methods to detect the intensity of social reaction with word-to-vector technique and context analysis. Tibebu *et al.* [9] analyzed real-time information on Twitter about opioid use and perceptions in Canada, which facilitated public health practice and opioid crisis addressing.

To explain and predict public emotional responses, especially the sentiment of distress and grief towards COVID-19, we analyze 999,978 microblogging posts from January 1, 2020 to February 18, 2020 on Sina Weibo[1] (Weibo for short), one of the most popular social media platforms in China with 550 million monthly active users in Quarter one, 2020. Weibo enjoys its traits of instant messaging, transparent sharing and publicly accessibility. In this work, the Deep Natural Language Processing (NLP) model and topic modelling

---

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.
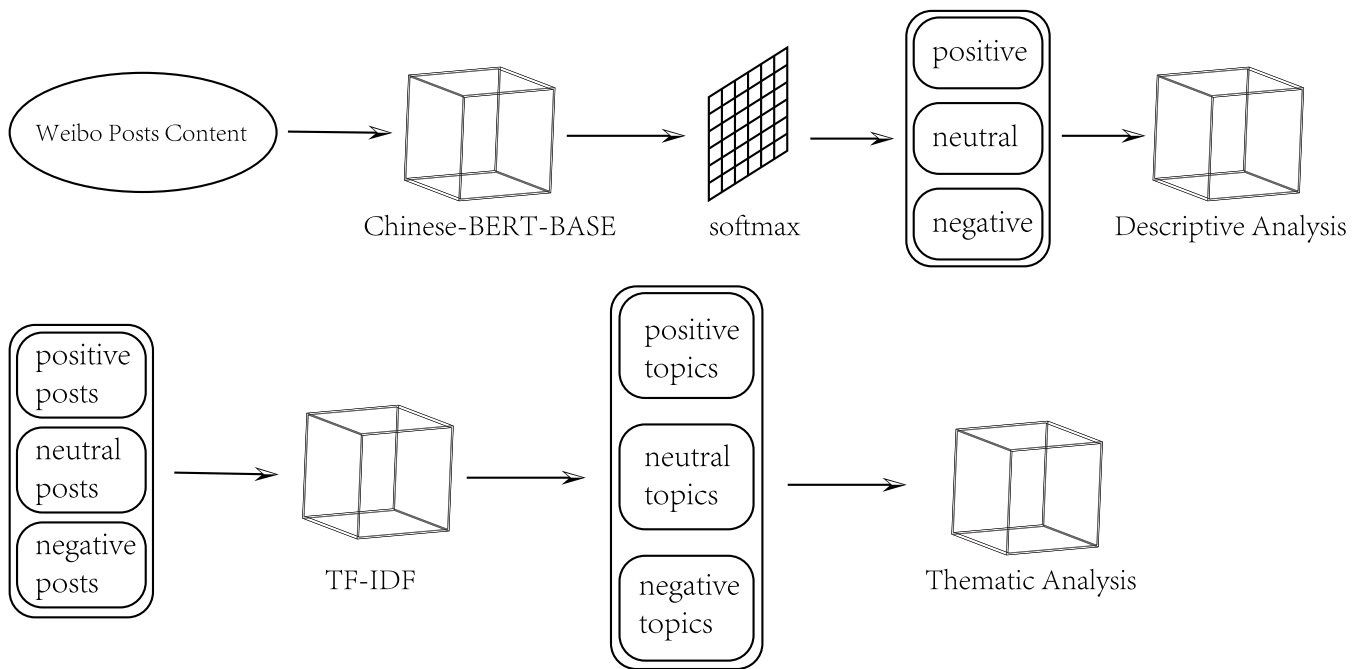
[1]https://www.weibo.com/

FIGURE 1. Workflow of sentiment analysis model.

method are utilized. To be specific, we fine-tune BERT for sentiment classification upon posts with three potential categories of sentiment, positive, neutral and negative, achieving a 75.65% of high accuracy, which surpasses many NLP baseline algorithms. The number of posts on each date is analyzed based on sentiment classification. Thereafter, TF-IDF model is adopted to extract central topics of posts. As the public sentiment on social media reflects people's psychological well-being and the spread of posts with negative sentiment may lead to social disruption and challenges for infection preventions [10], we analyze 11 dominant and distinctive topics extracted from Weibo posts with negative sentiment and investigate the trends of sentiment development and the underlying major themes to understand public concerns.

Outbreaks are now taking place in many countries around the world, especially Europe and North America [11]. For instance, as of 25 July 2020, there have been more than 4 million people affected in the U.S. Under this circumstance, main contributions of this study include:

- We fine-tune BERT model for sentiment classification on Chinese Weibo posts about COVID-19 and achieve considerable accuracy that beats all baseline NLP algorithms.
- The study demonstrates how public sentiment on social media evolves as COVID-19 spreads.
- We extract representative topics and discuss the dominant discourse of public distress about COVID-19 caused by related social events. Findings of this study could assist governments worldwide in making efficient and effective public health protection decisions.

II. RELATED WORK

Coronavirus Disease 2019 (COVID-19) is a newly occurred disease that related research has barely been published by the time of conducting our study. However, there has been some studies on text sentiment classification, which to some extent relate to our work. Ye *et al.* [12] applied Machine Learning SVM [13] model on Chinese product reviews for sentiment (positive or negative) classification and achieved better performance than the classical Semantic Orientation approach. Narayanan *et al.* [14] built a fast sentiment classifier using Naïve Bayes, which achieved 88.80% accuracy on popular IMDB movie reviews dataset. In recent years, researchers have used more deep learning neural network techniques on sentiment classification. Ren *et al.* [15] enhanced word representation with character embeddings and mainly applied CNN for a context-sensitive sentiment classification on Twitter contents. Tang *et al.* [16] proposed sentiment classification upon documents by a combination of LSTM/CNN embedding and gated RNN.

III. METHODOLOGY

Our sentiment analysis model consists of two parts, as shown in the workflow of Fig. 1. In particular, we first use fine-tuned BERT [17] to classify the sentiment of Weibo posts into positive, neutral and negative categories and analyze the trends of posts. Then we apply TF-IDF [18] algorithm to extract topics of posts with different sentiment. Specifically, 11 topics for negative posts are generalized, and we then analyze the underlying patterns.

A. SENTIMENT CLASSIFICATION AND DESCRIPTIVE ANALYSIS

Bidirectional Encoder Representations from Transformers (BERT), a neural network-based technique for natural language processing pre-training, has been largely applied in sentiment analysis [19]. The BERT model can be fine-tuned with proper input and output layers to create state-of-the-art models in a wide range of text analysis tasks [17]. The core of BERT is the adoption of transformer technique [20], which perfectly applies encoder-decoder model [21] and attention mechanism [22]–[24] on NLP tasks. In particular, attention mechanism allows the model to focus on the relevant parts of the input sequence as needed when input sequence is too long for typical NLP models to memorize all input features. In our work, attention mechanism in BERT model helps determine the relevance weight of each word token in input of Weibo posts and generate corresponding hidden states that can best describe the characteristics of different sentiment of Weibo posts during the training process by (1), where Q, K, and V are three vectors, query, key and value, created based on input embedding, and d_k is the dimension of key vectors.

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \qquad (1)$$

After sentiment classification, the percentage of each sentiment category is calculated. In public health crisis, people's responses evolve over time [25]. To gain the insights into people's reactions, Weibo posts with positive, neutral and negative sentiment of each day from 1 January 2020 to 18 February 2020 are compared. Furthermore, in order to identify dates with rapid changes in a number of posts with different sentiment categories, the increasing rate is calculated by (2) and further analyzed according to critical events happened on the corresponding dates.

$$\frac{\text{number of posts on a day} - \text{number of posts on the day before}}{\text{number of posts on the day before}} \qquad (2)$$

B. TOPIC EXTRACTION AND THEMATIC ANALYSIS

Term frequency-inverse document frequency (TF-IDF) is a numerical statistic reflecting how important a word is to a document in a collection or corpus [26]. The TF-IDF method is able to catch words that occur frequently by calculating term frequency and avoid insignificant words that occur in every document as important by the ability of inverse document frequency. Da Silva and Lopes [27] used TF-IDF to find the most informative Relevant Expression in each document in the corpus of their research. We perform the well-trained automatic sentiment classification model upon the unlabeled Weibo posts, and apply the built-in TF-IDF function in jieba [28], a Chinese word segmentation tool that operates based on its huge pre-trained corpus, to posts in each labeled sentiment

class for topic extraction by (3) for word t in a Weibo post d from the entire 999,978 Weibo posts dataset D.

$$\begin{cases} \text{tf}(t, d) = \log(1 + \text{freq}(t, d)) \\ \text{idf}(t, D) = \log(\frac{N}{\text{count}(d \in D : t \in d)}) \\ \text{tfidf}(t, d, D) = \text{tf}(t, d) \times \text{idf}(t, D) \end{cases} \qquad (3)$$

In detail, we treat each Weibo post as a document and each segmented Chinese token as a potential topic. Due to the fact that most Weibo posts are not too long (length distribution without an outlier 884 is as shown in Fig. 2), topics with top 5 TF-IDF scores in each post are extracted and analyzed. Topics in each sentiment category represent public focus and concerns regarding COVID-19. Further analysis based on the extracted topics is performed.
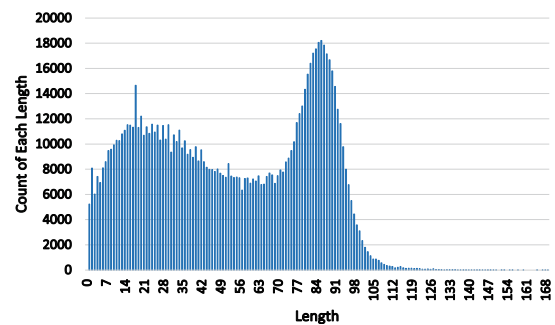

FIGURE 2. Post lengths of the whole Weibo post dataset. X-axis represents different post lengths, and y-axis represent the numbers of each post length.

Thematic analysis is a common tool to understand the perceptions and reasons for people's posts with negative sentiment [29]. In our case, topics appearing in more than 1% of total posts in each category of sentiment (positive, neutral and negative) are collected. Five topics are excluded, including Pneumonia (肺炎), Outbreak (疫情), Virus (病毒), Coronavirus (冠状病毒), and COVID-19 (新冠), because they directly represent COVID-19 and couldn't provide any valuable information in interpreting public sentiment. As a result, there are 38 key topics for positive sentiment, 19 for neutral sentiment, and 19 for negative sentiment, respectively. The 19 key topics for negative sentiment are compared with those of neutral sentiment and positive sentiment. Thereinto, 11 topics are found distinctive for negative sentiment and 8 topics are shared among posts with neutral or positive sentiment, where the latter are Masks (口罩), Wuhan (武汉), Definite Diagnosis (确诊), Doctor (医生), Case (病例), Infection (感染), Quarantine (隔离), and Hospital (医院). In addition, by exploring the semantics of topics, we divide the 11 negative-distinctive topics to four themes for further analysis. The four themes are Origin, Symptom, Production Activity, and Public Health Control. Moreover, we analyze the frequency of the 11 topics from 1 January 2020 to 18 February 2020 to visualize their trends.

IV. EXPERIMENTS AND RESULTS ANALYSIS

A. DATASET

Based on a list of 230 COVID-19 related key phrases, 2.4 million Weibo posts from 1 January 2020 to 19 February 2020 (posts on 19 February is incomplete thus excluded in our final sample) are crawled by CCIR 2020 organizer (26th China Conference on Information Retrieval) [30]. The crawler mainly uses SciPy and Beautiful Soup techniques, and deletion of duplicates and reposts are processed to construct the Weibo posts dataset. The dataset includes posts by around 640 thousand users with user location information excluded.

Manual sentiment labelling (positive, neutral and negative) is accomplished by CCIR 2020 organizer. In particular, open source Chinese sentiment analysis tools are adopted for preprocessing, and 12 volunteers were invited to finish the manual labeling on 120 thousand randomly selected Weibo posts from the dataset based on the preprocessing results and human decisions upon sentiment of Weibo contents. In detail, label of each Weibo post is decided by majority voting method of 3 volunteers who didn't know each other.

As a result, one million randomly selected posts from the whole dataset are shared to the public and 10% of them are manually labeled with three sentiment categories (positive, neutral and negative) by CCIR 2020 organizer.

B. SENTIMENT CLASSIFICATION

Manually labeled Weibo posts are randomly split into training and testing sets with a ratio of 5 to 5. In our experiment, we fine-tuned the Chinese BERT-BASE model with 12 layers and hidden dimension 768. According to the suggestion on hyperparameters selection by the original BERT paper [17], we set up parameters of 4 epochs, learning rate of 2e-5 and batch size of 32, and applied softmax neural network layer to train a three-category (positive, neutral and negative) sentiment classifier using the training dataset.

After training, the sentiment classification model achieves a 75.65% accuracy upon testing set. F1-scores on testing set along with precision and recall metrics for each sentiment category are summarized in Table 1. An overall weighted F1-score 0.7458 for classification model is obtained on data in Table 1 by (4), where w is the weight of each sentiment category. The sentiment classification is then performed upon the rest unlabeled Weibo posts.

$$\begin{cases} F_1 = \dfrac{2}{\text{recall}^{-1} + \text{precision}^{-1}} \\ \text{Weighted } F_1 = w_{\text{positive}} \times F_{1,\text{positive}} \\ \qquad\qquad + w_{\text{neutral}} \times F_{1,\text{neutral}} \\ \qquad\qquad + w_{\text{negative}} \times F_{1,\text{negative}} \end{cases} \quad (4)$$

We performed comparative tests on the same dataset using different baseline sentiment classification algorithms with the same data split ratios and random state values. Results are presented in Table 2.

TABLE 1. Performance measures of sentiment classification model.

	Precision(%)	Recall(%)	F1-score(%)
Positive	0.6990	0.7477	0.7225
Neutral	0.7975	0.7797	0.7885
Negative	0.6434	0.6246	0.6339

TABLE 2. Comparative test on NLP baseline algorithms.

System	Accuracy(%)
Fine-tuned BERT$_{\text{BASE}}$	75.65
SVM	70.66
Naïve Bayes	66.97
Logistic Regression	70.02
CNN	71.19
LSTM	57.73

TABLE 3. Examples of posts with different sentiment categories.

Sentiment	Post
Positive	#致敬白衣战士#辛苦了，感谢你们 #Salute to the warriors wearing white clothes (the doctor)# Thanks for your hard work. 中国加油，武汉加油，我为我的家乡加油 Stay strong China, stay strong Wuhan, I wish all the best for my hometown.
Neutral	世卫组织正式命名新冠肺炎为COVID-19 WHO officially name the new coronavirus as COVID-19. 我屋的窗户对着小区里的花园步道，还算空旷，平时经常有人跑步散步，今早正式通告了我家小区的确诊病例之后，一夜之间外面连个人影都很难看到了......? The view of my bedroom window is the often-spacious garden walkway of the neighborhood. There used to be people running or walking. However, after a confirmed case being announced in our community, it's hard to see even a single person.
Negative	好不容易要出院了然后我现在发烧38.5这是怎么了怎么了? Finally getting out of hospital and now I'm having a fever with 38.5°C. What's wrong with me, what's wrong with me? 过于无聊，复工时间又推后了，真担心哪天公司没了 So boring. Work Resumption has been postponed again. I'm so worried the company will be gone one day.

For 999,978 posts, most of them (56.2%) are neutral, while there are more positive sentiment (27.4%) than negative sentiment (16.4%). Examples of posts with different sentiment categories are given in Table 3. Specifically, posts of expressing gratitude would be labelled as positive while those about the fear for COVID-19 or life arrangements affected by COVID-19 would be categorized as negative. Neutral posts include the mere circulation of public information or plain descriptions of some life changes. From the upper part of Fig. 3, it is clear that COVID-19 related posts are increasing across the time and posts with different sentiment categories increase accordingly. Before 19 January, the number of posts about COVID-19 are quite stable. As to 20 January, there is an increase of 25.81% and 33.03% in negative posts and total posts, respectively, and a decrease of 16.30% in positive posts as compared to that of 19 January. Regarding the accumulative confirmed cases of COVID-19 in lower part of Fig. 3, it is clear that the number of posts in each day starts to grow from 20 January, in line with the increase in accumulative confirmed cases. The number remains relatively stable from early February during the early outbreak.
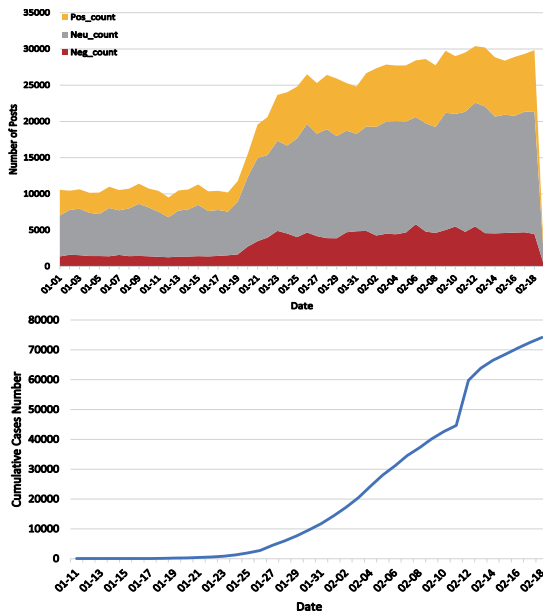
TABLE 4. 11 key topics in Weibo posts with negative sentiment along with the corresponding themes and frequency of occurrences in Weibo posts.

Key Topics	Themes	Percentage(%)
野味(Gamey Food)	Origin	3.08
蝙蝠(Bat)	Origin	2.70
阴谋论(Conspiracy Theory)	Origin	1.43
发烧(Fever)	Symptom	2.13
咳嗽(Cough)	Symptom	1.19
上班(Go to Work)	Production Activity	1.94
复工(Resume Work)	Production Activity	1.12
开学(School New Semester Beginning)	Production Activity	1.06
测体温(Temperature Taking)	Public Health Control	1.39
瞒报(Coronavirus Cover-up)	Public Health Control	1.26
封城(City Shutdown)	Public Health Control	1.09

The definitions of four themes in this work are as follows:
Origin: Discussions about where COVID-19 comes from.
Symptom: Change of body or mind indicating the possible infection of COVID-19.
Production Activity: Work life arrangements influenced by the outbreak of COVID-19.
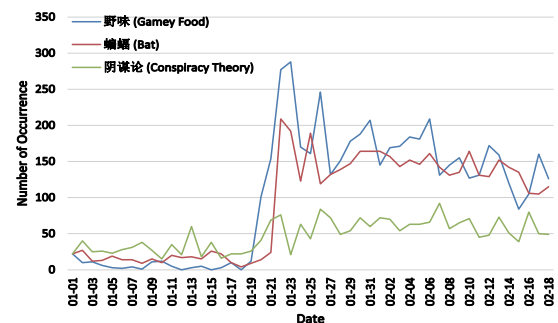Public Health Control: Measures to stop the spread of COVID-19.

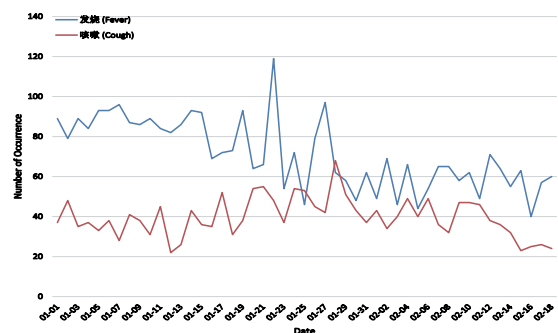**FIGURE 4.** Number of posts related to the Origin with respect to the dates.

FIGURE 5. Number of posts related to the Symptom with respect to the dates.

COVID-19 is first noticed in China in December 2019 [1], but the number of Weibo posts remains stable and relatively low during the early period. On 20 January 2020, there is a surge in total Weibo posts and posts with negative sentiment as well, and the number of posts is kept at that level ever since. One important event might be related to this. On the night of 20 January Dr. Zhong Nanshan confirmed the human-to-human transmission of COVID-19 on China Central Television [31]. This surge of Weibo posts indicates the tremendous influence of information revealed by government-related media and it is therefore important for government to discreetly handle the public depression about COVID-19.

C. NEGATIVE SENTIMENT ANALYSIS

As listed in Table 4, the 11 key topics of Weibo posts with negative sentiment fall into four themes: Origin, Symptom, Production Activity, and Public Health Control. Typically, the number of posts in each theme for each day from 1 January 2020 to 18 February 2020 are plotted in Fig. 4-7 for further analysis.

There are three topics about the Origin of COVID-19 as shown in Fig. 4. In particular, "Gamey Food" (3.08%) and "Bat" (2.70%) are two primary assumptions for the origins of COVID-19 [5], [32]. They remain low frequency in the early January. However, there is a sharp increase in posts of "Gamey Food" on 20 and "Bat" on 22, January 2020, both then keeping a high frequency thereafter and reaching the highest number of posts with negative sentiment per day among 11 topics. And the "Conspiracy Theory" (1.43%) suggesting that COVID-19 does not have a natural origin is condemned by scientists [33] but widespread on social media. However, it remains in a relatively low frequency in

comparison with "Gamey Food" and "Bat". The discussions of the origins of virus are deeply correlated to negative sentiment and trigger the largest amount of posts with negative sentiment as demonstrated in Fig. 4. However, as discussed, rumors and unconfirmed information may overwhelm the discussion [33]. Consequently, it is important for the government to release the transparent progress about the investigation of origins.

"Fever" (2.13%) and "Cough" (1.19%) are identified as the representative symptoms of COVID-19 [34]. For posts about Symptom as shown in Fig. 5, posts with "Fever"

as topic outnumbered that with "Cough", and the gap is quite large in early January but declines in February as COVID-19 spread. Symptoms about COVID-19 might lead to negative sentiment but could also be beneficial for self-detection of infection. Therefore, typical symptoms of disease should be revealed to the public clearly and timely, which would benefit early detection of the disease.

Production Activity as pictured in Fig. 6 summarizes topics about work-life arrangements under the threat of COVID-19. "Go to Work" (1.94%) and "Resume Work" (1.12%) portray the public concern over work, and "School New Semester Beginning" (1.06%) indicates people's worries about students going back to school. The concern on "Go to Work" starts from early January and remains relatively high while the one on "School New Semester Beginning" starts to grow from 20 January 2020 and the worries for "Resume work" starts to grow from 26 January 2020. The arrangements for production activity might be an important driving force for the general depression. The peak of concern for "School New Semester Beginning" in Fig. 6 comes earlier than that for "Resume Work", thus earlier arrangements are necessary in easing people's tension.
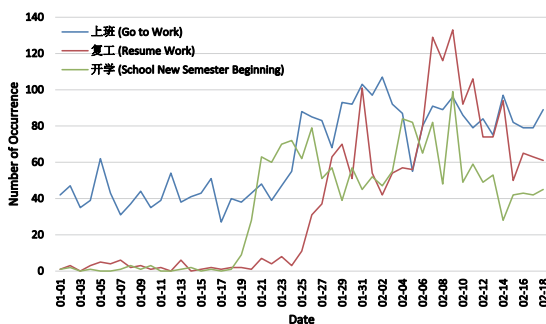

FIGURE 6. Number of posts related to the Production Activity with respect to the dates.
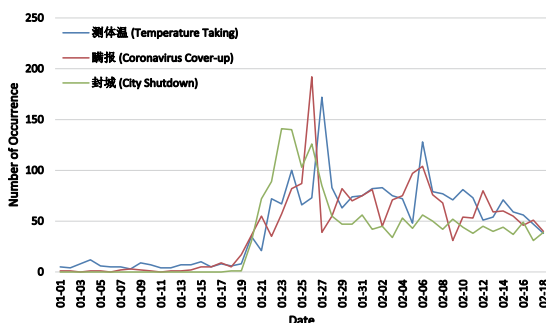

FIGURE 7. Number of posts related to the Public Health Control with respect to the dates.

Three topics about Public Health Control (Fig. 7) depict different aspects. "Temperature Taking" (1.39%) is an important method for COVID-19 diagnosis. "Coronavirus Cover-up" (1.26%) is the misbehavior or fault in public health control which might lead to risk of more infections [35]. And "City Shutdown" (1.09%) describes the strict public health control taken by the Chinese government [36].

There are almost no negative posts about Public Health Control in the early January. The peak discussions are "City Shutdown" the first on January 23, "Coronavirus Cover-up" the second on January 26, and "Temperature taking" the third on January 27, and the three topics reach the same level of attention around 18 February 2020. This trend demonstrates the focuses on different control measures as situation develops. Towards this end, the public health control and opacity of information may also lead to people's depression. Therefore, the corresponding public health control measures at a certain time period should be elaborated for public acceptance.

V. CONCLUSION

The spread of COVID-19 has turned to a worldwide pandemic thus far [37]. Public health concerns not only relate to the infection prevention but also the psychological status of people experiencing the disaster [38]. Therefore, analyzing posts with negative sentiment from social media could contribute to understanding the experiences of Chinese general public during the outbreak of COVID-19 and offers examples for other countries. Our analyses provide insights on the evolution of social sentiment over time and the topic themes connected to negative sentiment of Weibo posts. Fig. 3 illustrates the clear outbreak dates for public attention about COVID-19. Moreover, concerns about Origin, Symptom, Production Activity, and Public Health Control are deeply intertwined with the public sentiment.

This study collects data on social media from early stage of COVID-19 transmission in China. Based on the data analysis and discussion, several advantages emerge. First, state-of-the-art fine-tuned BERT classification model and TF-IDF topic extraction model deliver results with considerable accuracy. Second, it can further be implemented as an online platform for real-time monitoring on public sentiment during other crises in the future. Third, this study reveals important topic themes which are deeply connected to sentiment of depression. As the infection of COVID-19 keeps spreading worldwide now, insights from this study may contribute to public administration and prevention of social disruptions.

Despite of informative results found in this study, further improvements are expected on the classification model to achieve a higher accuracy. Furthermore, only information on Sina Weibo is used in this study, which may lead to bias by neglecting posts on other social media platforms. Finally, in order to focus on the centrality of topics, only topics appearing in more than 1% of total posts are selected in each category of sentiment. This may lead to the overlook of important topics with less percentage. Future studies by incorporating information in empirical data from different social media platforms and different countries may contribute to a more solid conclusion.

New outbreaks are taking places in many other countries all around the world. The sentiment classification model and findings of this study would provide constructive instructions for governments worldwide on making efficient and effective public health protection decisions.

ACKNOWLEDGMENT

REFERENCES

[1] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of COVID-19 is higher compared to SARS coronavirus," *J. Travel Med.*, vol. 27, no. 2, pp. 1–4, Mar. 2020, doi: 10.1093/jtm/taaa021.

[2] *Novel Coronavirus (2019-nCoV) Situation Report—1*. World Health Organization. Accessed: Mar. 22, 2020. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4

[3] S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang, D. He, and M. H. Wang, "Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," *Int. J. Infectious Diseases*, vol. 92, pp. 214–217, Mar. 2020.

[4] *Coronavirus Disease 2019 (COVID-19) Situation Report—45*. World Health Organization. Accessed: Mar. 22, 2020. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200305-sitrep-45-covid-19.pdf?sfvrsn=ed2ba78b_4

[5] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *Int. J. Antimicrobial Agents*, vol. 55, no. 3, Mar. 2020, Art. no. 105924.

[6] Z. Li *et al.*, "Vicarious traumatization in the general public, members, and non-members of medical teams aiding in COVID-19 control," *Brain, Behav., Immunity*, vol. 88, pp. 916–919, Aug. 2020, doi: 10.1016/j.bbi.2020.03.007.

[7] X. Gui, Y. Kou, K. H. Pine, and Y. Chen, "Managing uncertainty: Using social media for risk assessment during a public health crisis," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2017, pp. 4520–4533.

[8] J. Pei, G. Yu, X. Tian, and M. R. Donnelley, "A new method for early detection of mass concern about public health issues," *J. Risk Res.*, vol. 20, no. 4, pp. 516–532, Apr. 2017.

[9] S. Tibebu, V. C. Chang, C.-A. Drouin, W. Thompson, and M. T. Do, "At-a-glance–What can social media tell us about the opioid crisis in canada?" *Health Promotion Chronic Disease Prevention Canada*, vol. 38, no. 6, pp. 263–267, Jun. 2018.

[10] X. Ji, S. A. Chun, Z. Wei, and J. Geller, "Twitter sentiment classification for measuring public health concerns," *Social Netw. Anal. Mining*, vol. 5, no. 1, p. 13, Dec. 2015, doi: 10.1007/s13278-015-0253-5.

[11] *World Map*. Centers for Disease Control and Prevention. Accessed: Mar. 22, 2020. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/world-map.html

[12] Q. Ye, B. Lin, and Y.-J. Li, "Sentiment classification for chinese reviews: A comparison between SVM and semantic approaches," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Aug. 2005, pp. 2341–2346, doi: 10.1109/ICMLC.2005.1527335.

[13] V. V. Naumovic, "SVM and logistic regression," in *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 2000, pp. 156–163.

[14] V. Narayanan, I. Arora, and A. Bhatia, "Fast and accurate sentiment classification using an enhanced naive Bayes model," in *Intelligent Data Engineering and Automated Learning—IDEAL* (Lecture Notes in Computer Science), vol. 8206. Berlin, Germany: Springer, 2013, pp. 194–201.

[15] Y. Ren, Y. Zhang, M. Zhang, and D. Ji, "Context-sensitive twitter sentiment classification using neural network," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 215–221.

[16] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1422–1432, doi: 10.18653/v1/D15-1167.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language Understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

[18] C. Sammut and G. I. Webb, "TF-IDF," in *Encyclopedia of Machine Learning*. Boston, MA, USA: Springer, 2011, pp. 986–987.

[19] Z. Gao, A. Feng, X. Song, and X. Wu, "Target-dependent sentiment classification with BERT," *IEEE Access*, vol. 7, pp. 154290–154299, 2019.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 6000–6010.

[21] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 2, 2014, pp. 3104–3112.

[22] T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1412–1421.

[23] D. D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," presented at the 6th Int. Conf. Learn. Represent., Vancouver, BC, Canada, Apr. 2018 [Online]. Available: https://arxiv.org/abs/1409.0473

[24] Y. Kim, C. Denton, L. Hoang, and A. M. Rush, "Structured attention networks," presented at the 5th Int. Conf. Learn. Represent., Palais des Congrès Neptune, Toulon, France, Apr. 2017. [Online]. Available: https://arxiv.org/abs/1702.00887

[25] B. T. Burkholder and M. J. Toole, "Evolution of complex disasters," *Lancet*, vol. 346, no. 8981, pp. 1012–1015, Oct. 1995.

[26] A. Rajaraman and J. D. Ullman, "Data mining," in *Mining Massive Datasets*. Cambridge, U.K.: Cambridge Univ. Press, 2011, pp. 1–17.

[27] J. F. D. Silva and G. P. Lopes, "A document descriptor extractor based on relevant expressions," i n *Progress in Artificial Intelligence (EPIA)*. vol. 5816. Berlin, Germany: Springer, 2009, pp. 646–657.

[28] *Jieba*. PyPI. Accessed: Mar. 22, 2020. [Online]. Available: https://pypi.org/project/jieba

[29] L. Mollema, I. A. Harmsen, E. Broekhuizen, R. Clijnk, H. De Melker, T. Paulussen, G. Kok, R. Ruiter, and E. Das, "Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013," *J. Med. Internet Res.*, vol. 17, no. 5, p. e128, May 2015.

[30] *CCIR 2020*. CCIR2020. Accessed: Mar. 22, 2020. [Online]. Available: http://www.cvnis.net/ccir2020/index.html

[31] Huanqiu.com. *Dr. Zhong Nanshan Confirmed the Human-to-Human Transmission of COVID-19*. Accessed: Mar. 22, 2020. [Online]. Available: https://china.huanqiu.com/article/9CaKrnKoZPT

[32] Z. Hou, L. Lin, L. Lu, F. Du, M. Qian, Y. Liang, J. Zhang, and H. Yu, "Public exposure to live animals, behavioural change, and support in containment measures in response to COVID-19 outbreak: A population-based cross sectional survey in China," *medRxiv Preprints*, Feb. 2020, doi: 10.1101/2020.02.21.20026146.

[33] C. Calisher *et al.*, "Statement in support of the scientists, public health professionals, and medical professionals of China combatting COVID-19," *Lancet*, vol. 395, no. 10226, pp. e42–e43, Mar. 2020.

[34] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao, H. Liu, L. Zhu, Y. Tai, C. Bai, T. Gao, J. Song, P. Xia, J. Dong, J. Zhao, and F.-S. Wang, "Pathological findings of COVID-19 associated with acute respiratory distress syndrome," *Lancet Respiratory Med.*, vol. 8, no. 4, pp. 420–422, Apr. 2020.

[35] Z. Xu, S. Li, S. Tian, H. Li, and L.-Q. Kong, "Full spectrum of COVID-19 severity still being depicted," *Lancet*, vol. 395, no. 10228, pp. 947–948, Mar. 2020.

[36] Y. Xiao and M. E. Torok, "Taking the right measures to control COVID-19," *Lancet Infectious Diseases*, vol. 20, no. 5, pp. 523–524, May 2020.

[37] *Coronavirus Disease (COVID-19)—Events as They Happen*. World Health Organization. Accessed: Mar. 22, 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen

[38] S. C. Vos and M. M. Buckner, "Social media messages in an emerging health crisis: Tweeting bird flu," *J. Health Commun.*, vol. 21, no. 3, pp. 301–308, Mar. 2016.

TIANYI WANG is currently pursuing the Ph.D. degree with the Department of Computer Science, The University of Hong Kong. His major research interests include machine learning, deep learning, and cyber security.

KE LU is currently pursuing the Ph.D. degree with the Department of Social Work and Social Administration, The University of Hong Kong. His major research interests include nonprofits, organization development, social enterprise, social innovation, crowdfunding, and philanthropy. Recently, he has been involved in studying the organization development of education nonprofits in China.

QING ZHU received the Ph.D. degree in cardiology from Shandong University, Jinan, China, in 2003.

She is currently a Cardiologist and a Chief Physician with the Qilu Hospital of Shandong University, and also an Associate Professor and the Master Tutor of the Medical College of Shandong University. Her research interests include the clinical prevention treatment and basic research of lipid metabolism abnormality and atherosclerosis, electrophysiological mechanism of arrhythmia and radiofrequency ablation, and chronic disease management of cardiovascular disease.

● ● ●

KAM PUI CHOW received the Ph.D. degree in computer science from the University of California, Santa Barbara, USA.

He is currently an Associate Professor with the Department of Computer Science, and the Director of the Center for Information Security and Cryptography (CISC), The University of Hong Kong (HKU). He is also the Leader of the Computer Forensics Research Group (CFRG). In the years 1994–1997, he together with other professionals and a team of software engineers developed the search engine for Hong Kong Telecom's 108 Telephone Directory Enquiry System using state of the art technology in main memory database and distributed computing. In the recent years, his research interests include migrated to digital forensics and computer security. His research interests include computer forensics, digital investigation, data privacy, cryptography, and computer security.