Journal of Engg. Research Vol. 9 No. (3A) September 2021 pp. 219-239 10.36909/jer.v9i3A.8703

Tracking patients healthcare experiences during the COVID-19 outbreak: Topic modeling and sentiment analysis of doctor reviews

Adnan M. Shah*, Xiangbin Yan, Samia tariq and Syed Asad A. Shah

Department of Management Science and Engineering, School of Management, Harbin Institute of Technology, Harbin, 150001, P.R China
**Corresponding Author: adnanshah486@gmail.com*

Submitted: 27/09/2019
Revised: 12/12/2020
Accepted: 19/12/2020

ABSTRACT

Emerging voices of patients in the form of opinions and expectations about the quality of care can improve healthcare service quality. A large volume of patients' opinions as online doctor reviews (ODRs) are available online to access, analyze, and improve patients' perceptions. This paper aims to explore COVID-19-related conversations, complaints, and sentiments using ODRs posted by users of the physician rating website. We analyzed 96,234 ODRs of 5,621 physicians from a prominent health rating website in the United Kingdom (*Iwantgreatcare.org*) in three-time slices (i.e., from February 01 to October 31, 2020). We employed machine learning approach, dynamic topic modeling, to identify prominent bigrams, salient topics and labels, sentiments embedded in reviews and topics, and patient-perceived root cause and strengths, weaknesses, opportunities, and threats (SWOT) analyses to examine SWOT for healthcare organizations. This method finds a total of 30 latent topics with 10 topics across each time slice. The current study identified new discussion topics about COVID-19 occurring from time slice 1 to time slice 3, such as news about the COVID-19 pandemic, violence against the lockdown, quarantine process and quarantine centers at different locations, and vaccine development/treatment to stop virus spread. Sentiment analysis reveals that fear for novel pathogen prevails across all topics. Based on the SWOT analysis, our findings provide a clue for doctors, hospitals, and government officials to enhance patients' satisfaction and minimize dissatisfaction by satisfying their needs and improve the quality of care during the COVID-19 crisis.

Keywords: COVID-19; Text mining; Topic modeling; Doctor reviews; Sentiment analysis.

INTRODUCTION

Coronavirus pandemic (COVID-19) is a fast-spreading pathogen that is primarily transmitted through breathing and close contact and is normally infectious to humans. COVID-19 has been declared a global pandemic by the World Health Organization (WHO) (WHO, 2020a). The pandemic, which emerged from Wuhan, China, has spread in more than 200 countries, with 49,242,837 confirmed cases and cost 1,242,187 lives by November 8, 2020 (WHO, 2020c). The United Kingdom (UK) occupies the largest proportion of confirmed cases (1,171,445) and deaths (48,888), as of November 8, 2020 (WHO, 2020d). Without vaccine availability in the market, health authorities across the UK are seeking to monitor transmission and reduce pressures on the healthcare system through lockdown, school and business closures along with staying at home orders. There is a need for careful strategic planning and management to reduce disease risk and minimize the long-term economic effects.

To date, despite extensive clinical studies on different diseases, no vaccine for COVID-19 has been developed; only supporting strategies are being used for those who need critical care. Public prevention measures remain essential to slowing its spread. Social distance and, in some cases, social isolation have been one of the most successful strategies to halt or avoid the spread of COVID-19 (Janz & Becker, 1984; WHO, 2020b). Furthermore, hygienic practices, such as frequent hand washing and wearing face masks, can reduce the propagation of the pathogen (WHO, 2020b). Such easy but effective steps require robust communications, displayed on widely viewed communication channels, to be disseminated effectively. Moreover, with the rapidly changing pandemic situation, the pace of information transmission has become very critical.

One of the quickest and most approachable communication channels for broadcasting information is social media (Ali *et al.*, 2020). Social media have grown drastically in recent years. An increasing number of healthcare officials and public healthcare departments are using social networking channels for collaboration and exchange of knowledge during crises. Social media have become a critical platform for encouraging risk communication during public health emergencies (Househ, 2016; Gui *et al.*, 2018). Tracking public discussions on social media regarding health and policy subjects provides a barometer of the Britain's and global opinion of COVID19. This is mainly important because the circumstances with COVID-19 shift every day and are impulsive during these volatile periods.

Social networking sites have been used as a notifier for early warnings, an emergency coordination channel, a public opinion tracking, and a public health monitoring database for a number of disasters and outbreaks (Zhao *et al.*, 2020). Because of the worsening situation in the UK, discussions on social media about the pandemic have been radically rising since March 2020. Social media played a critical role before the pathogen spread and continues to do so as it diffuses internationally. As a result, social networking sites have become a lifeline for the public to access information, share opinions, and socialize during crisis (Ordun *et al.*, 2020b).

The current work helps expand the broader array of literature on social media analytics and COVID-19. The COVID-19 pathogen requires not only quick dissemination of information, but also information recognition and arrangement. In particular, user-generated content in social media can be extracted to assess the public's information, patterns, and responses toward the diseases (Li *et al.*, 2020). However, social networking sites and other internet platforms are not regulated by the Government bodies (Zhao *et al.*, 2020). It has been found that these sites and platforms contain a high volume of rumors and misinformation regarding the COVID-19 pandemic that has become a challenging task for major social media outlets (Convertino, 2020). In the past, several attempts were made to analyze social media data to measure public perceptions regarding the COVID-19 disease outbreak (Lwin *et al.*, 2020; Zhao *et al.*, 2020; Dong *et al.*, 2020; Chen *et al.*, 2020a; Xue *et al.*, 2020a; Chen *et al.*, 2020b; Li *et al.*, 2020; Zhu *et al.*, 2020; Yin *et al.*, 2020; Jang *et al.*, 2020; Ordun *et al.*, 2020a; Saleh *et al.*, 2020; Hung *et al.*, 2020; Stokes *et al.*, 2020). Our research can further add to this literature by showing that online data mining can provide speedy insights into public perceptions and opinions.

This paper is distinct from other research in that we examined datasets of online doctor reviews (ODRs) from a UK-based digital platform (iWantGreatCare, a subsidiary of National Health Service). This digital platform is controlled by healthcare agencies in the UK. A systemic tracking process exists whenever people write online reviews about the quality of treatment received from doctors. The content of ODRs is monitored by PRW owners. An ODR that includes abusive material, misinformation, misrepresentations, and rumors cannot be published on the physician rating websites. These contents are immediately removed by the platform filtering mechanism before they are available to the public.

This study used social media analytics to analyze the ODRs dataset related to COVID-19. The significant discussion topics in ODRs are captured by dynamic topic modeling (DTM). We also assessed trends of hot topics of discussion and emotional tendencies related to the COVID-19 crisis over three COVID-19 cycles across 96,234 COVID-19-related comments on a PRW. Furthermore, we investigate the sentiment and trends of ODRs about COVID19. Finally, we examined the identified topics to explore the medical system's strengths and weaknesses during the COVID-19 crisis. The current research uses a massive quantity of collected ODRs to respond to our perception of the pandemic and contribute knowledge to it.

This study examines the trends in topics and mood during three different time slices of pandemic using a hybrid computational approach. The specific research questions in this study include the following:

1) *What are the latent topics of concern during the COVID-19 outbreak*?
2) *How the topics and their moods do change over time*?
3) *How do PRWs users' sentiments expressed in ODRs change over time*?
4) *How do the identified topics and sentiments guide healthcare providers and officials in improving patient satisfaction toward the medical system?*

Investigating various topics and related emotions will show the shifting nature of public experience during the crisis. The current study hypothesized that performing topic mining and sentiment analysis of ODRs during the different time slices of the COVID-19 pandemic could provide useful insight into the public's views and perceptions on this policy. Our work also demonstrates how Natural Language Processing (NLP) techniques could be extended to patients posted ODRs. This study's results will also shed light on unobserved feelings and patterns associated with the COVID-19 pandemic. To the best of our knowledge, this is the first attempt to identify different topics and sentiment trends across different time-periods of COVID-19 pandemic. The knowledge mining may be useful for communication on public health and for disseminating and refining information strategies to perform strengths, weaknesses, opportunities, and threats (SWOT) analyses for strategic planning in hospitals.

LITERATURE REVIEW

The focus of the current research includes social media analytics during the COVID-19 outbreak and topic modeling. As social media provides us the public perceptions regarding the COVID-19 pandemic, we review some popular studies on analyzing textual content from social media platforms.

With the outspread of COVID-19 pathogen globally, scholars have suggested various text mining techniques based on social media data for tracking public perceptions regarding the outbreak. Topic mining and sentiment analysis approaches to machine learning are common strategies for analyzing textual content on social media. Doogan *et al.* (2020) proposed a MetaLDA topic model to identify patterns that affected public opinions and attitudes about non-pharmaceutical interventions (for instance, lockdown and travel restrictions) during the early stage of the COVID-19 crisis. Zhu *et al.* (2020) applied an LDA topic model to identify the critical topics of concern and trends in topics and emotions changing with time during the COVID-19 outbreak. Kim (2020) analyzed news big data regarding the COVID-19 outbreak using the topic modeling approach. Jelodar *et al.* (2020) used an LDA topic modeling to identify critical issues associated with COVID-19 from people's opinions.

Moreover, sentiment analysis was performed to identify public opinion trends using LSTM recurrent neural network. Yin *et al.* (2020) proposed a new method to detect important topics and sentiment dynamics due to the COVID-19 crisis from social media posts. Jang *et al.* (2020) analyzed Twitter data regarding the COVID-19 outbreak using topic modeling and aspect-based sentiment analysis. Moreover, the current study analyzed topic trends changes with the time during the pandemic. Ordun *et al.* (2020a) used topic modeling to generate different topics that discuss different aspects of COVID-19 pandemic. Zhao *et al.* (2020) analyzed the topic trends and sentiment related to the COVID-19 crisis. Chen *et al.* (2020b) used topic modeling to discover different controversial terms related to different topics and sentiments. Xue *et al.* (2020a) used topic modeling analysis to identify topics and sentiments from tweets. Furthermore, the study analyzed the change in sentiment trends over time. Stokes *et al.* (2020) used a topic modeling approach to identify different topics and measure daily trends in discussion topics on an online public forum during the COVID-19 outbreak. Hung *et al.* (2020) used topic modeling to analyze different discussions regarding COVID-19 using Twitter data and examined the sentiments toward this novel virus. Gozzi *et al.* (2020) employed the topic modeling approach to define media coverage and collective online response toward the COVID-19 outbreak. Saleh *et al.* (2020) examined the bunches of conversation using topic modeling and related sentiments.

The textual comments related to the COVID-19 posted in social media can be deemed potentially beneficial to extract important topics to better figure out public views and perception and strengthen health strategies. Although extant research mainly focused on social media to analyze people's opinions regarding COVID-19 in the Chinese and

United States context, to the best of the author's knowledge this is the first study to employ NLP techniques to analyze ODRs regarding COVID-19 from a high traffic ranking PRW in the UK. This site is being controlled by National Health Service UK. A systemic monitoring process exists on this rating site when people post online reviews about the quality of treatment received from doctors. The ODRs are regulated by PRW operators. An ODR with abusive material, misinformation, false statements, and rumors cannot be published on PRW. The filtering mechanism for the website automatically deletes such content before it is available to the public.

Based on the DTM approach and sentiment analysis, the current study automatically mine hot topics and classify these topics and comments to understand the people's positive or negative perceptions about the COVID-19 crisis to make appropriate decisions.

PROPOSED METHODOLOGY

Research design

The current social media analytics research included data preparation and data analytics steps. Data preparation involved three steps: (1) sampling; (2) data collection; and (3) raw data preprocessing. After cleaning the raw dataset, we moved to the data analytics phase, including (1) Using DTM techniques to identify hot topics with their corresponding bigrams related to different time-frame of the COVID-19; (2) analyzing topic trends and their sentiments using NLP techniques; (3) discrete emotions in Plutchik's framework of emotions was computed using the algorithm *CrystalFeel*, a sentiment analysis tool; (4) SWOT analysis and root cause analysis is performed using the topic theme-based sentiment summary (TSS). The unit of analysis for this research was each ODR posted on PRW.

Sample and data collection

We purposely selected *Iwantgreatcare.org* as our research context. This rating website is an affiliated enterprise of National Health Service, a UK-based health corporation offering more than 1 million physicians' information. We used a network spider programmed in Python version 3.6.1 software to collect ODRs published between February 01, 2020, and October 31, 2020; a total of 98,331 reviews of 5798 physicians were collected. After removing the non-English and duplicate ODRs, 96,234 ODRs were included in our dataset for this study. Figure 1 describes the methodology provided in the current study.

Data preprocessing

The raw data was cleaned using Python programming language in the following steps:

1) We corrected typographical errors in the text.
2) We transformed all reviews to low-case text.
3) Using the Python package NLP toolkit (Bird *et al.*, 2009), we removed numbers, special characters, URLs, punctuations, stop words, symbols, redundant words, prepositions, and pronouns. These contents were not useful for data analysis.
4) We removed those words with occurrence lesser than 10 times in the corpus.
5) We converted all repeated word charterers its base form. For example, veryyyy nice was converted to very nice.
6) We removed all white spaces from review sentences to generate tokens when necessary.
7) We tokenized the refined text into a bag of words.
8) Next, we used the part-of-speech (POS) tagger to tag each word (particularly nouns and adjectives) in a sentence (Mihalcea & Tarau, 2004).
9) Finally, we stemmed the words to their standardized base form using a snowball algorithm (Porter, 2001).

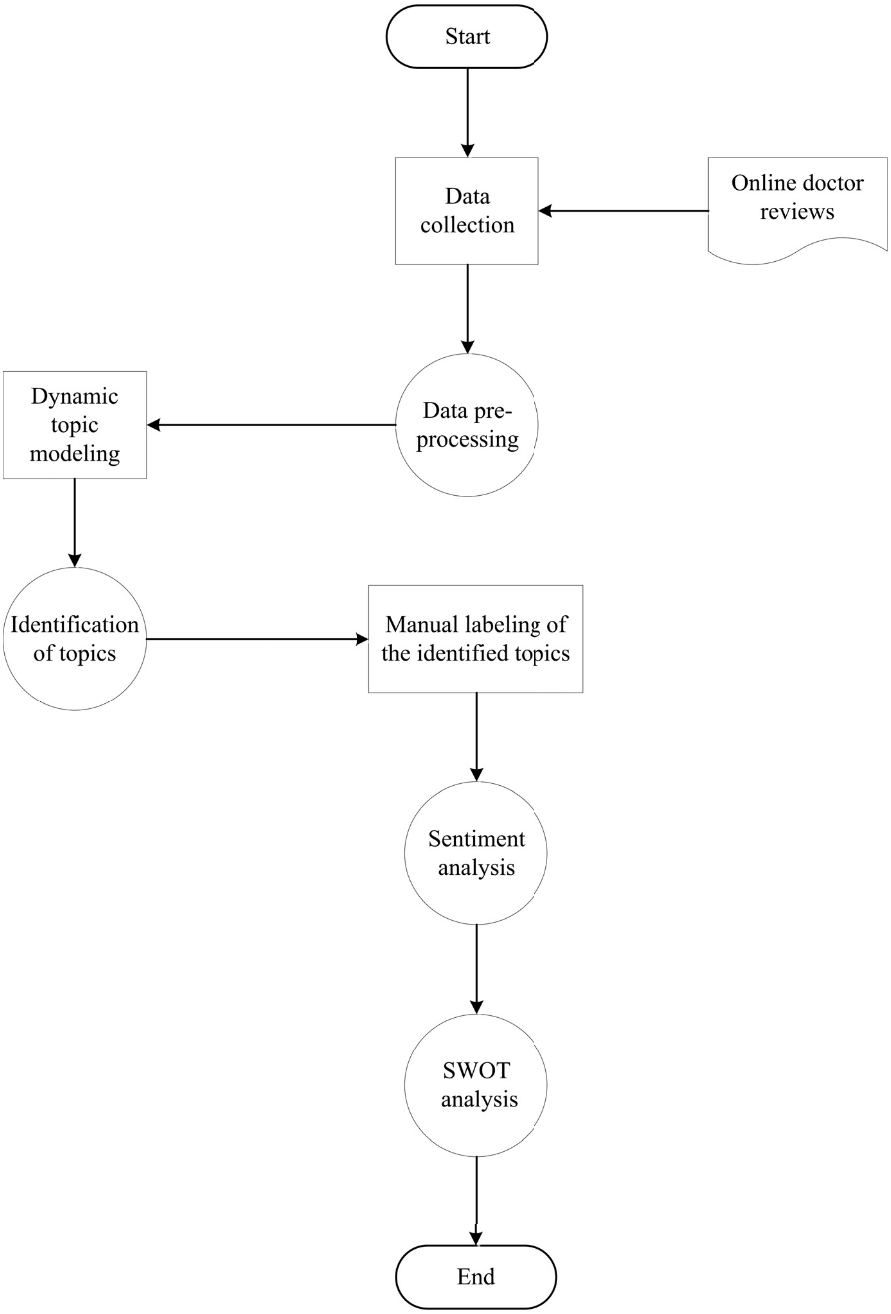

Figure 1. Overall summary diagram of the proposed approach.

Data analysis

Dynamic topic modeling

Topic modeling (Boyd-Graber *et al.*, 2017) is used to identify hidden (latent) semantic structures that associate documents by defining the frequently occurring keywords. Ranked keywords associated with topics are collective illustrations of the documents in which they present. Various algorithms may be used for the topic modeling. Topic modeling, initially proposed by Blei *et al.* (2003) has been frequently used. Several extensions of the LDA model have been developed in previous studies (Abdellaoui *et al.*, 2018; Doogan *et al.*, 2020; Sha *et al.*, 2020; Ordun *et al.*, 2020a; Paul & Dredze, 2014, 2011).

We used dynamic topic models (Jiaxiang *et al.*, 2020; Gensim, 2019; Luke, 2018), a generative model that enables the discovery of topic trends over time (Gerrish & Blei, 2010; Blei & Lafferty, 2006). Various scholars used DTM to explore public opinions in online communities (Cao & Tang, 2014; Parra *et al.*, 2016; Ha *et al.*, 2017). In the DTM, each document contained a number of topics and words that pursue a multinomial distribution. DTM can also manage sequential data and create topics for different time slices of the corpus. In particular, the documents are modeled on the static LDA in each slice, and topics relevant to the slice t develop from the topics from the earlier slice t - 1. DTM extracts topics of each time slice with LDA and their parameters α and β are connected into a state space model that evolves with Gaussian noise to bring a smooth development of the topics from time to time.

Model parameters are estimated through the variational approach employing the variational Kalman Filtering and the variational wavelet regression. Gensim (Gensim, 2019) and NLTK (*http://www.nltk.org*), python libraries for NLP, were used to apply the DTM analysis to the current study. The number of topics is a critical parameter in topic modeling and allowing these topics humanly interpretable (Stevens *et al.*, 2012). As in (Blei & Lafferty, 2006), considering the first slice of DTM, we employ LDA on the first day of the dataset to learn the most suitable number of topics. We employ the Gensim package (Rehurek, 2014) in Python to train the LDA model to select the best topic number. In addition, a coherence score was used to calculate the degree of semantic similarity between high scoring topic words as a signal for selecting the suitable number of topics. In other words, the word's co-occurrence relationship is expressed in vectors, and semantic similarity is the cosine similarity of word vectors. The coherence score is the average of these similarities (Röder *et al.*, 2015). The coherence score helps to differentiate between human interpretable topics and statistical measures (Grimmer & Stewart, 2017).

$$Coherence = \sum_{i<j} score(w_i, w_j) \tag{1}$$

We chose top n keywords words with the highest frequency in each topic and then accumulate all the top n keywords pairwise scores $w_i,\ldots,w_n$. Figure 2 presents the coherence values of different topic numbers using the LDA model based on one day ODRs. We can observe that the maximum value of the coherence score is 0.67 as the number of topics reached to 30. We conclude that the total number of topics is constant and fix the DTM topic numbers as 30. DTM algorithm identified ten topics during each time slice and addressed relevant matters related to them.

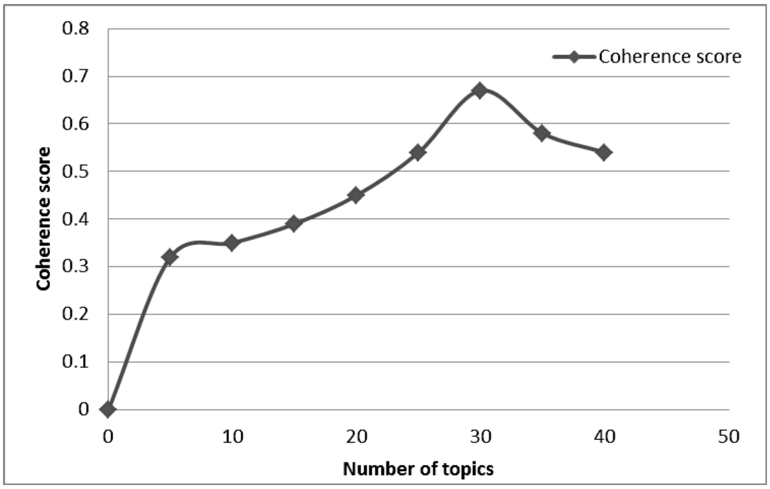

Figure 2. Number of topics selected for coherence score.

Qualitative analysis

After identifying the most popular topics from ODRs using DTM, we employed the qualitative approach to support deeper qualitative dives into the dataset, such as assigning labels to the bigrams, interpreting the topics and patterns identified from the ODRs generated by machine learning algorithms. The qualitative method depends on the diverse and detailed understanding from humans, which enables theoretical approaches to be inductive, exploratory, and applicable.

Sentiment analysis

Sentiment analysis is a computational technique often generally referred to as opinion mining that attempts to evaluate the thoughts, emotions, and behaviors of people in a given text (Yin *et al.*, 2020). Sentiment analysis has been widely used and an important tool in social media analytics research (Yin *et al.*, 2020; Xue *et al.*, 2020a; Lwin *et al.*, 2020; Colnerič & Demšar, 2020). Recently, sentiment analysis techniques are very advanced, and many software are openly available, such as Stanford's CoreNLP, CrystalFeel, VADER, SenticNet, SentiStrength, and SentiCircles. The sentiment analysis in the current study was based on *CrystalFeel* algorithm, a sentiment analytics tool whose accuracy had been well-established (Gupta & Yang, 2018). This model categorized each emotion into four discrete emotions in Plutchik's framework of emotions (Plutchik, 2001), including joy, surprise, and trust (positive), and anger, fear, sadness, and disgust (negative). This approach returned one emotion from the eight classifications for each given ODR.

RESULTS AND DISCUSSION

After the data collection for three time periods, 98,331 reviews were retrieved, representing 5798 within the four geographic areas of the UK, i.e., England, Northern Ireland, Scotland, and Wales. Figure 3 shows the number of ODRs related to COVID-19 from February 01, 2020, and October 31, 2020. A peak of the ODRs volume is identified between March 11 and 15 and then gradually increase again after March 15, 2020 (see Figure 3). This result is also timed by the WHO declaration COVID-19 as a global pandemic. A growing volume of ODRs may be followed by the WHO declaration COVID-19 as a global pandemic, which gives a strong opportunity to direct the public to take precautionary action in March. The number of ODRs gradually increased since March 10, 2020 and dropped after May 27, 2020, due to strict lockdown, travel restrictions, and a decrease in new cases. However, posting ODRs increased since September 13, 2020, due to a sudden increase in the number of new cases as a part of the second wave of the COVID-19 and news about the development of the vaccine. The speed at which ODRs can be found posted on PRWs may lead the public and officials to react to the virus spread in the early phases.

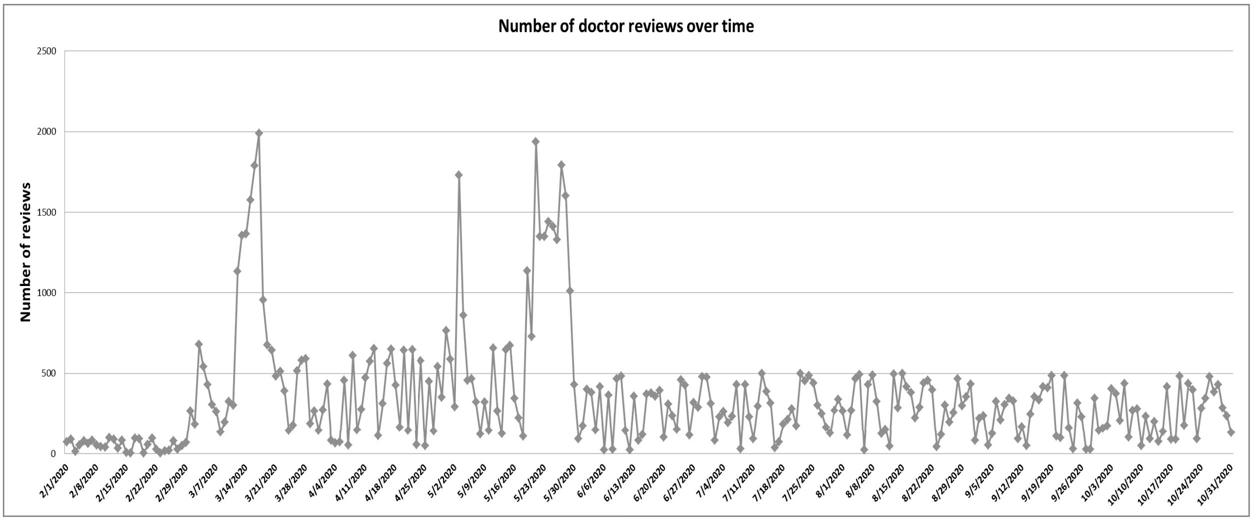

Figure 3. The number of doctor reviews over time.

Topics identification related to COVID-19

To discover topics and track the topic change over time, we construct topic models on our PRW data using DTM implement in the Gensim package (Gensim, 2019) and NLP toolkit in Python. Keeping in view the coherence score of 30, 10 topics were identified for each time slice. For the dataset used in the current study, we selected the number of topics 30 extracted by DTM since it had the highest coherence score. The top ten extracted topics for each time slice with the most important keywords (bigrams) related to the topics are shown in Table 1. Two authors discussed the bigrams in each of the 10 topics for each time slice and then labeled them as a topic.

Table 1. Topics and their corresponding contributed bigrams across different time slices.

	Feb 01, 2020 – April 30, 2020	May 01, 2020 – July 31, 2020	Aug 01, 2020 – Oct 31, 2020
Topic 1	**Response toward disease**	**Hospital business process**	**Everyday situation**
0	People response	Hospital facilities	Stores open
1	Ignore measures	Medical staff	Essential available
2	Die early	Social distancing	Company hiring
3	Get rid	Healthcare	Businesses growth
4	Seriously involve	Patients appointment	Told everyone
5	Disease spread	Health care	Grocery prices
6	Take seriously	Public attention	Jobs lost
7	Say no	City center	Paid workers
8	Go hospital	Health workers	Restaurants open
9	Intelligence reports	Person-person contact	Customers demand
Topic 2	**Reported cases**	**Face mask**	**Economic impact**
0	Cases reported	Wonder about	Store timings
1	New cases	Stay home	Stay home
2	Death toll	Expected supply	Business closed
3	Important instructions	Sell mask	Home stay
4	Total deaths	Wear mask	Sick workers
5	Report cases	High price	Working capacity
6	Food stuff	Make mask	Workers’ wages
7	Bank loans	N-95 mask	Employees
8	Covid-19	Face cover	Job lost
9	Far areas	Show concern	Weeks longer
Topic 3	**Stay home**	**Health care**	**Social responsibility**
0	Safe place	Health outcome	Counties contributions
1	Home quarantine	Drink water	Donations money
2	Stay safe	Right direction	Federal control
3	Panic situation	Company policies	Pandemic control
4	Request public	Regular exercise	United combat
5	Complete procedures	Risk reduction	Pound collection
6	Healthy activities	Result positive	British government
7	Toxic relationship	Healthy diet	Hoax minimization
8	Employee vacations	Scarce resources	National cause

9	Go early	Would appreciate	Million pounds
Topic 4	**Violence against lock down**	**Stop spread**	**COVID-19 consequences**
0	Rubber bullets	Slow transmission	Biological war
1	Black masks	Deadly virus	Global crisis
2	Public benefit	Save citizen	Buying power
3	Water cannon	Act responsibly	Parts
4	COVID pandemic	Move forward	Goods prices
5	Road traffic	Stop rumors	Panic buying
6	Die hunger	Woman responsibility	Run business
7	Entire day	Leader vision	Chain outlets
8	Business closed	Stop spread	Stocks availability
9	People injured	Virus transmission	Manufacturing
Topic 5	**Travel restrictions**	**Social distancing**	**Gratitude healthcare and reduce spread**
0	Airports diversion	Thing settle	National cause
1	Boarder restrictions	Think sharply	Stay united
2	Bus stands	Month checkup	Stay safe
3	Long-distance	Finally	Get rid
4	Spread fast	Get rid	Maintain strength
5	Longer suspension	Article publish	Human cause
6	Flight operation	Get frustrated	Donate more
7	Transport closed	Vaccine development	Washing your hands
8	Temporarily suspension	Go with	Remember poor
9	Strict monitoring	Social restrictions	Follow advices
Topic 6	**COVID-19 news**	**Work**	**Activities**
0	Want know	Day tip	Exercise daily
1	Health emergency	Need assistance	Diet equilibrium
2	London markets	Difficult time	Yoga classes
3	Breaking news	Street life	Coffee cup
4	Self-isolate	Lead front	Online teaching
5	Good idea	Little effort	Weekend movie
6	News web	Create space	Run slow
7	Health services	Online medium	Nearby park
8	Health officials	Know more	Home schooling
9	Fake news	Work regularly	School canceled
Topic 7	**Administration prompt response**	**Quarantine process and outcomes**	**Life during pandemic**
0	Big move	Test positive	Fine health
1	NHS policies	Number patients	Hard time
2	Testing procedures	Follow instructions	Months long
3	Prompt response	Virus spread	Live tough
4	WHO declaration	Death toll	Started business

5	Infectious disease	Spread slow	Safe journey
6	Disease management	Report new	Good
7	Health authorities	State emergency	Hope good
8	Ventilators shortage	Country policy	Feel relax
9	Testing kits	Strict lockdown	House rent
Topic 8	**Virus Symptoms**	**Safety during pandemic**	**Second wave of COVID**
0	Body aches	Sick people	Confirmed cases
1	Abdominal discomfort	Great effort	Death rate
2	Stuffy nose	Spread slow	Mortality rate
3	Breath shortness	Start business	Public health
4	Severe headache	Person satisfaction	Toll rises
5	Sore throat	Told everyone	Climate change
6	Skin rash	Contact less	Medical staff
7	More fatigue	Family safe	Tested positive
8	High fever	Diffuse slow	Realize intensity
9	Frequent cough	Becoming normal	Full swing
Topic 9	**Health policies**	**Virus transmission**	**Vaccine development/ treatment**
0	Stay safe	Emergency situation	Phase III
1	Corona virus	Novel pneumonia	Researcher experiments
2	Stop spread	Isolation centers	Big relief
3	Health minister	First-level response	Genetic problem
4	Health crisis	Goods and materials	Launch December
5	Good thing	Nationwide lockdown	Early trials
6	God bless	Confirmed diagnosis	Partner companies
7	Case scenario	Wear masks	First launch
8	Infectious disease	Startup opportunity	Combat virus
9	Worst case	First case	Clinical trials
Topic 10	**Pathogen spread**	**Life normalcy**	**Supply chain during COVID-19**
0	Confirmed cases	Adopt situation	Gloves protection
1	Bringing total	Pandemic normal	Business growth
2	Birmingham city	Talk about	Sanitizer shortage
3	Cases confirmed	Hard time	Control mechanism
4	Number confirmed	Good news	Materials shortage
5	New confirmed	Life normal	Production capacity
6	Coronavirus cases	Ray hope	Increasing demand
7	Total number	Good time	Supply shortage
8	Cases reported	Time spent	Manufacture cost
9	New deaths	Forgot failures	Taxes exemption
Note: Topic labels in bold			

Topics trends across three different time slices

We analyze the trends in health-related topics over different time slices. We performed a fundamental analysis focused on an evaluation of the model's estimates of the document-to-topic distribution. We first divide ODRs into three different time slices, e.g., Feb 01–April 30, May 01–July 31, Aug 01–Oct 31. Next, we measure a mean vector θ for ODRs for each time slice as presented in (Jang *et al.*, 2020). Referring to the mean vector θ for each time slice, we plot graphs of health-related topics over time, as presented in Figure 4-6.

For time slice 1, we can see that the trends in the three-time slices are extremely consistent and homogeneous. While minor variations exist, the overall rise and decline trends are approximately uniform. For instance, the topics regarding violence against lockdown (T4) and travel restrictions (T5) reach their highest level from mid-February till mid-April and then significantly decrease. During time slice 2, the topic patterns are very much linked to the public health initiatives. For time slice 2, people started to concentrate more on topics facemask (T2), healthcare (T3), social distancing (T5), work (T6), quarantine process and outcomes (T7), safety during pandemic (T8), virus transmission (T9). The topic about work (T6) shows an uneven behavior; therefore, this topic starts to increase slightly till May 10, then decreases till May 20, then again starts to increase during the entire time slice 2. For time slice 3, interestingly, economic impact (T2) starts to gradually increase till August 10, then decline after that till the entire time slice. This is because the British government announced a massive aid package to the business community belong to the areas hit hardest by the COVID-19. Therefore, the economic impact due to the COVID-19 becomes a less popular discussion issue after August 10. Next, the topic regarding vaccine development/treatment (T9) has been the hot topic of discussion during the entire time slice 3. This is because the British healthcare officials have signed a COVID-19 vaccine contract to purchase approximately 60 million doses of trial treatment from multi-national medicine giants GSK and Sanofi. The healthcare officials have also signed a deal for 100 million vaccine doses with the Oxford University being produced by AstraZeneca. Officials have also ensured 90 million more doses for two other effective vaccines.

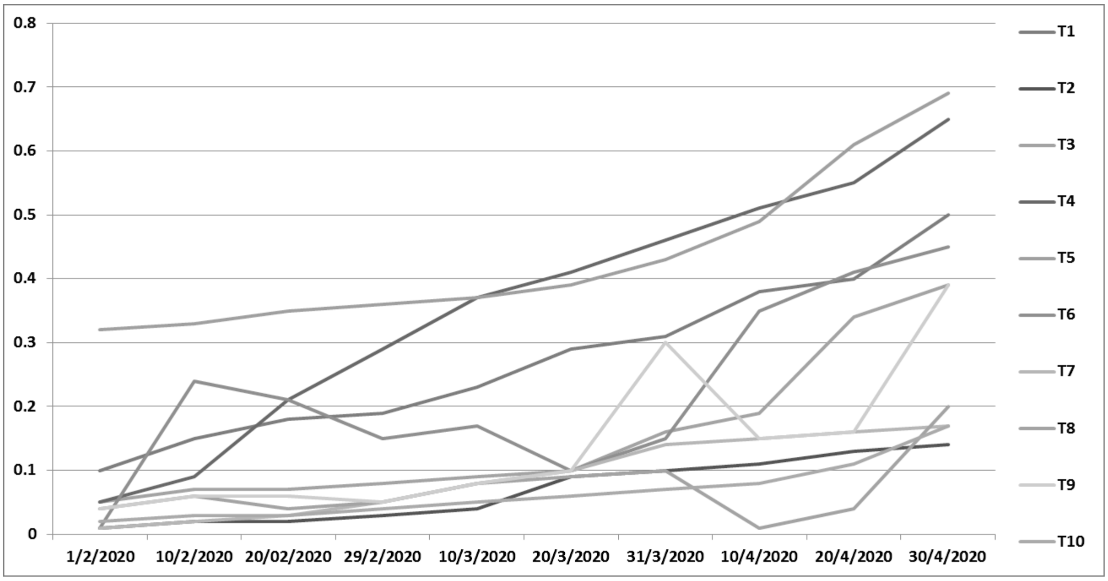

Figure 4. Visualization of topic trends for time slice 1.

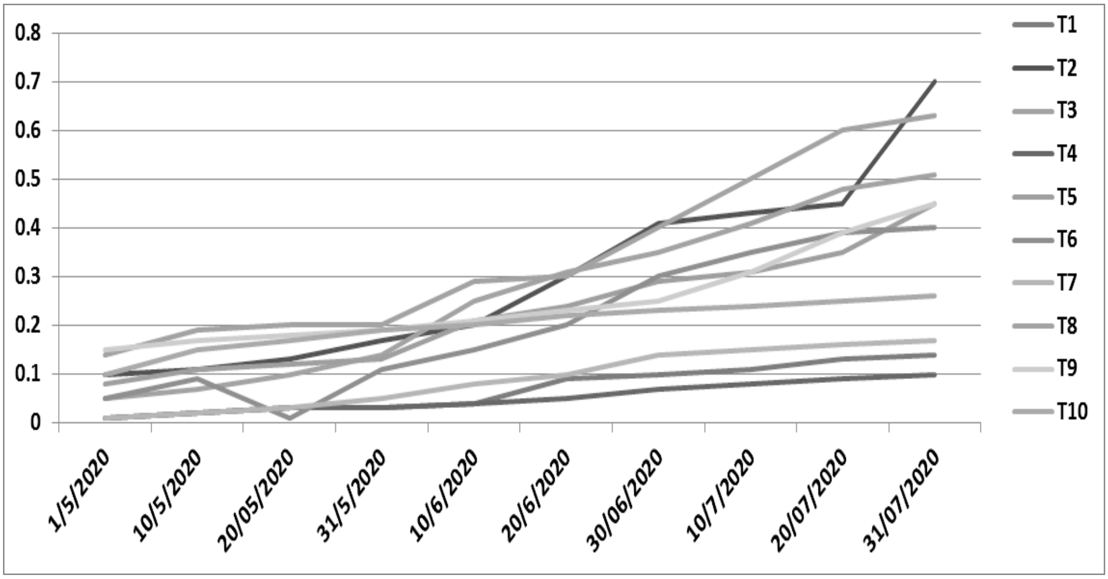

Figure 5. Visualization of topic trends for time slice 2.

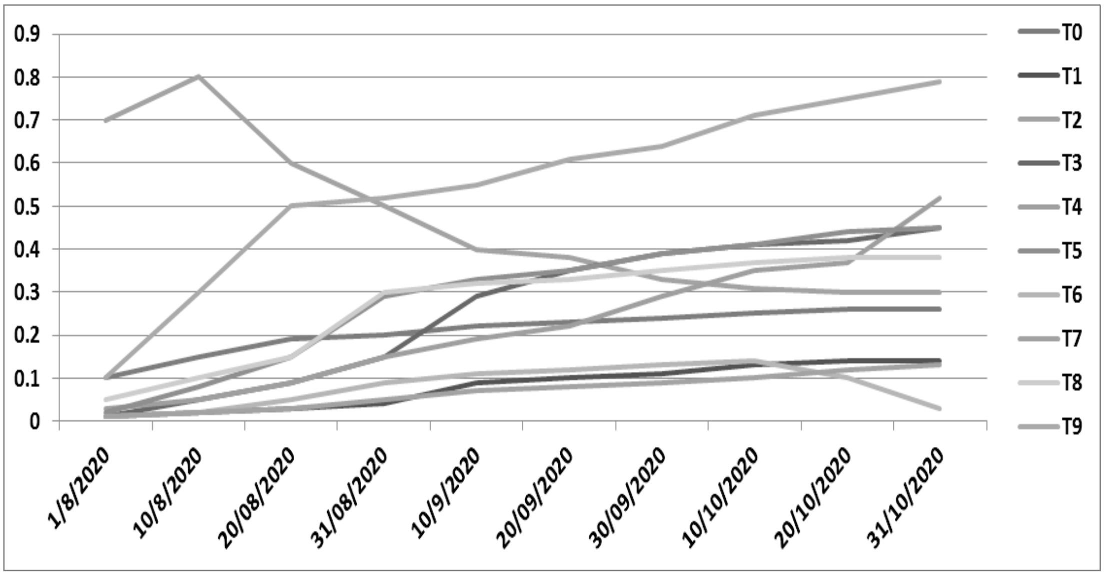

Figure 6. Visualization of topic trends for time slice 3.

Online doctor reviews sentiments analysis

The ODRs provided data about the thoughts and feelings of the public. The emotional response of the public towards the COVID-19 crisis is shown in Figure 7. It reflected the percentage of emotional ODRs over daily ODRs (with ten days difference) by date across three-time slices.

Fear was the most prevailing emotion during the first time slice when the pandemic first appeared in the UK with hundreds of new cases, which was about 55% of daily ODRs. The proliferation of fear falls below 35% of daily ODRs in the second time slice due to ease in travel restrictions and lockdown. In contrast, ODRs on anger progressively increased from early-March to mid-May, peaking at 27% on March 12, a day after the WHO announcement about the COVID-19 as a global pandemic. Since then, ODRs about anger declined significantly but remained at a reasonably high level. Sadness related ODRs have been doubled since the WHO announcement but are relatively smaller than those of the other emotions. Likewise, ODRs on joy also increased, indicating a sense of superiority, appreciation, expectation, and pleasure.

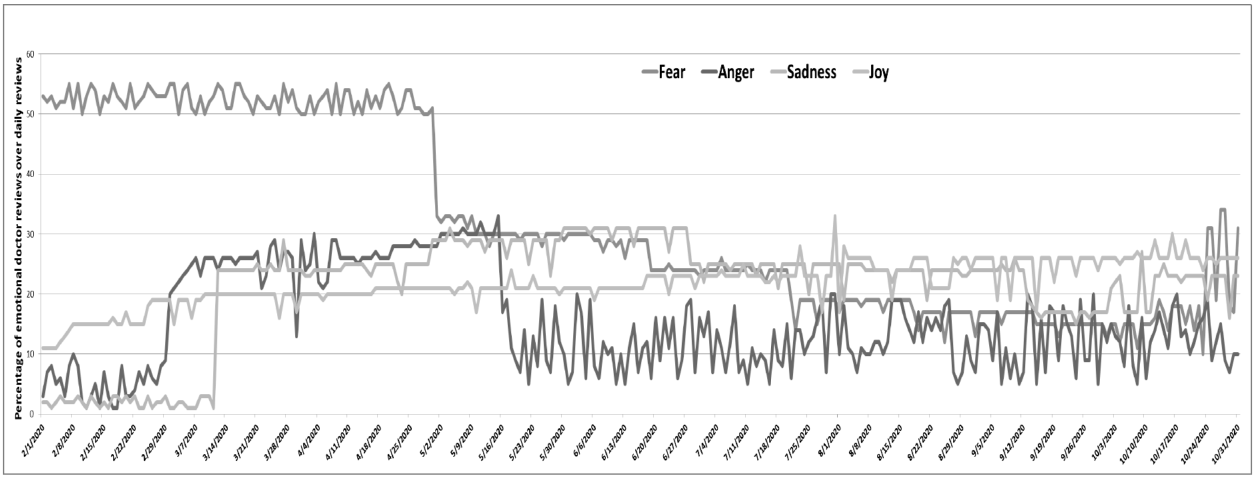

Figure 7. Emotional trends during the COVID-19 pandemic.

Measuring sentiments within the extracted topics

In this section, all the ODRs are clustered into their respective topics and labeled with emotion polarity. The opinion polarity is combined with ODRs to measure the overall distribution of emotion. For each topic across each time slice, we add up the number of positive and negative ODRs in the topic, such that each topic is linked with two emotion counts.

Table 2 and Figure 8 displayed the proportion of each emotion contained by each of the 10 topics for each time slice. We found that the feeling of fear was dominant across all topics. For instance, fear of the uncertain circumstances due to the COVID-19 was about 47% of the ODRs in all thirty topics. About 22% of ODRs' emotions under Topic 3 linked to the public's confidence in health officials.

Since fear was prevalent in all 30 topics, we further ran a one-tailed *z* test and analyzed the statistically relevant differences in each of the eight emotions across topics. A *p-value* less than .001 was used as a threshold and provided findings in Table 2.

For instance, fear of COVID-19 uncertainty has been found to be more likely to dominate in Topics 1, 2, 4, 5, 6, and 10 [for time slice 1], Topics 2, 4, 5, 7, and 8, [for time slice 2], Topics 2, 4, and 7 [for time slice 3]. Trust communicated in ODRs was statistically significant in the Topics 3 and 9 [for time slice 1], Topics 1, 3, and 6 [for time slice 2], Topics 3, 6, and 10 [for time slice 3]. Surprise regarding the COVID-19 crisis was statistically significant most discussed in Topic 8 [for time slice 1], Topic 9 [for time slice 2], and Topic 8 [for time slice 3]. Finally, Joy as the component of discrete emotions has been statistically significant in Topic 7 [for time slice 1], Topic10 [for time slice 2], and Topics 1, 5, and 9 [for time slice 3].

Table 2. The percentage of 8 emotions across 30 topics.

Topic	Anger, %	Anticipation, %	Disgust, %	Fear, %	Joy, %	Sadness, %	Surprise, %	Trust, %
1	1.21	1.21[b]	0.3[b]	45.31[b]	15.11[b]	1.51[b]	8.98[b]	30.31[b]
2	1.32[b]	1.51[b]	0.3[b]	45.42[b]	14.12[b]	1.67[b]	7.65	28.65[b]
3	1.42	1.51[b]	0.3[b]	45.62[b]	12.34[b]	1.74[b]	5.45[b]	25.65
4	1.51[b]	1.50	0.4[b]	44.21	13.54[b]	1.8[b]	4.44[b]	23.11[b]
5	1.62[b]	1.55	0.31[b]	46.10	14.56	1.21	3.44	24.67
6	1.63[b]	1.41[b]	0.31	45.40[b]	12.56[b]	1.35[b]	4.56[b]	25.63[b]
7	1.65[b]	1.51[b]	0.61[b]	47.15	20.11[b]	1.54[b]	5.67	23.11[b]
8	1.66[b]	1.62	0.31[b]	47.22[b]	18.54[b]	1.71	5.44[b]	22.14[b]
9	1.68[b]	1.51[b]	0.72	47.18	14.32	1.41[b]	3.45[b]	22.15[b]
10	1.69[b]	1.52[b]	0.32[b]	47.42[b]	11.23[b]	1.78[b]	6.54[b]	22.17[b]
11	1.51	1.81	0.82[b]	47.16	12.23	1.87[b]	6.76	23.45[b]
12	1.52	1.54	0.34[b]	47.12[b]	14.43[b]	1.98	9.78[b]	24.65
13	1.53[b]	1.51[b]	0.32	47.90	15.16	1.23	9.87[b]	25.78
14	1.54	1.51[b]	0.41[b]	47.15[b]	16.34[b]	1.24[b]	9.56[b]	26.19[b]
15	1.55	2.10	0.31[b]	47.19	12.45	1.25[b]	9.67[b]	27.88[b]
16	1.30[b]	1.52[b]	0.53	47.54[b]	15.43[b]	1.27[b]	5.67	29.09
17	1.57[b]	1.52[b]	0.31[b]	45.12	17.15	1.65	5.78[b]	30.11[b]
18	1.56[b]	1.50	0.32[b]	44.15[b]	19.65	1.43[b]	4.45[b]	25.22[b]
19	1.10	2.41[b]	0.65	48.11	17.55[b]	1.44[b]	3.65[b]	26.72
20	1.51	2.21[b]	0.37[b]	48.09[b]	15.32	1.43[b]	3.64[b]	26.65[b]
21	1.57[b]	2.31	0.37[b]	48.07	14.34	1.23[b]	7.6[b]	26.12
22	1.73[b]	1.51[b]	0.51[b]	45.12[b]	13.22[b]	1.24	7.66	27.86[b]
23	1.76	1.52[b]	0.32[b]	45.13[b]	12.23	1.98[b]	7.56[b]	28.23
24	1.71[b]	1.10	0.30	45.15[b]	13.34[b]	1.97[b]	8.76	29.17[b]
25	1.22	1.55	0.71[b]	45.19[b]	14.45[b]	1.95[b]	6.66[b]	30.33
26	1.73[b]	1.56	0.31	46.11	15.55	1.65[b]	8.87[b]	22.32[b]
27	1.74[b]	1.31[b]	0.90[b]	47.09[b]	15.66[b]	1.76	8.81[b]	21.12[b]
28	1.31	1.59[b]	0.33[b]	47.32[b]	15.41[b]	1.65[b]	8.99	28.54[b]
29	1.76[b]	1.26[b]	0.41	45.67[b]	17.71	1.44[b]	8.01[b]	26.17[b]
30	1.73[b]	1.30[b]	0.35[b]	48.11[b]	18.91[b]	1.34[b]	5.32[b]	29.12[b]

[a] The sum of the proportion for each topic is not equal to 100%. The rest are composed of either neutral or other emotions.
[b] $p<.001$ from z test.

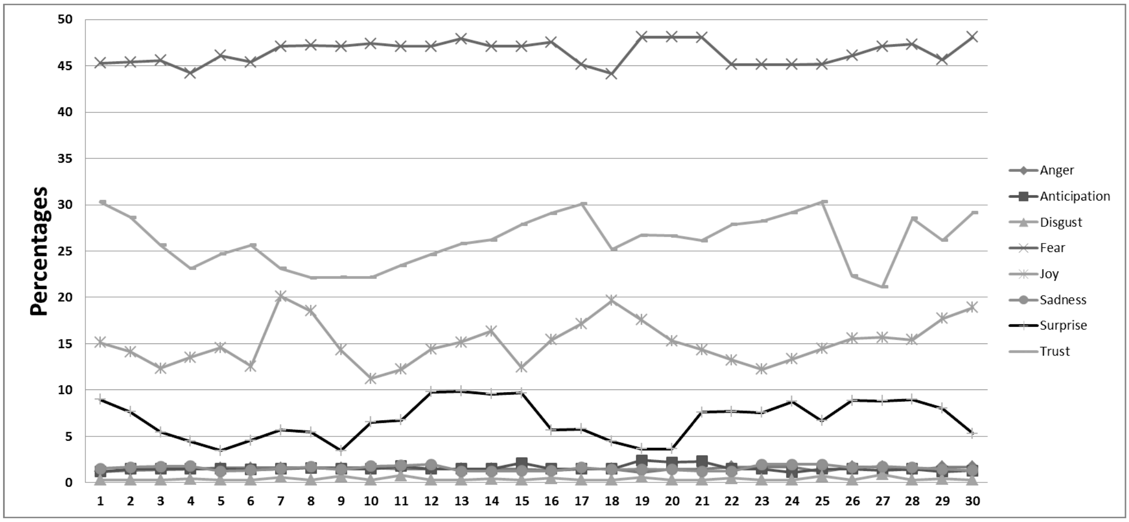

Figure 8. Sentiment analysis for each of 30 topics.

SWOT analysis

For each physician, the opinion polarity score of all the review sentences is added together by their topics to retrieve the TSS. The findings help perform SWOT analysis for the health care system under review. SWOT is a technique most widely used to identify the strategic components (internal and external) critical to business growth. The strengths and weaknesses are defined by analyzing the enterprise's internal features, whereas opportunities and threats are external elements like competition. The fundamental structure of SWOT analysis is presented in Figure 9. For instance, SWOT analysis underlines the key components that affect patients' perspectives on their treatment experience and can help the healthcare system plan decisions for the health care system in their decision-making.

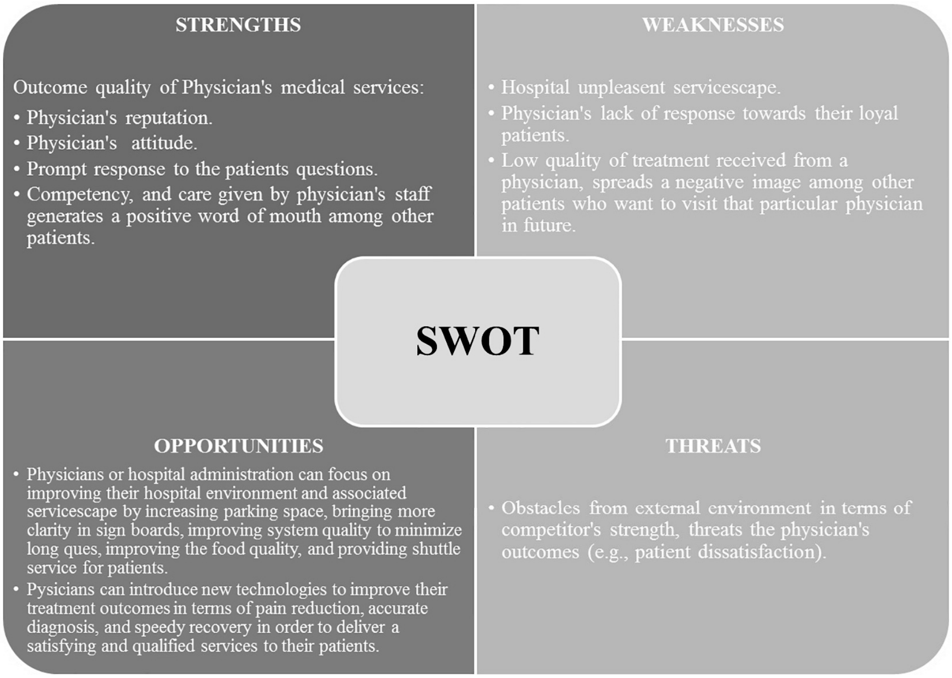

Figure 9. The basic structure of SWOT analysis.

Figure 10 presents the TSS in medical service quality settings, emphasizing the overall positive (strengths) and negative (weaknesses) topics from the patient's viewpoint. Furthermore, the majority of the doctor reviews express negative on the following topics: 'Response toward disease,' 'Reported cases,' 'Violence against lock down,' 'Travel restrictions,' 'COVID-19 news,' and 'Pathogen spread' [for time slice 1]; 'Face mask,' 'Stop spread,' 'Social distancing,' 'Quarantine process and outcomes,' and 'Safety during pandemic' [for time slice 2]; and 'Economic impact,' 'COVID-19 consequences,' and 'Life during pandemic'[for time slice 3]. On the other hand, patient spreads a positive word of mouth among other patients regarding the following topics: 'Stay home,' 'Health policies,' 'Administration prompt response,' and 'Virus Symptoms' [for time slice 1]; 'Hospital business process,' 'Health care,' 'Work,' 'Virus transmission,' and 'Life normalcy' [for time slice 2]; and 'Everyday situation,' 'Social responsibility,' 'Gratitude healthcare and reduce spread,' 'Activities,' 'Second wave of COVID,' 'Vaccine development/treatment,' and 'Supply chain during the COVID-19' [for time slice 3].

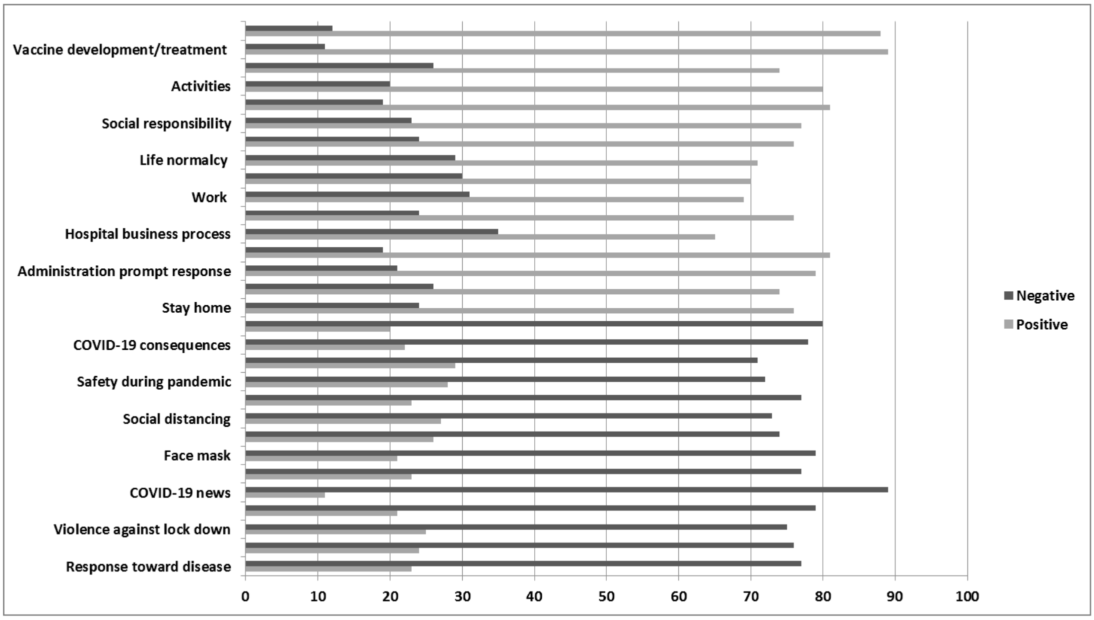

Figure 10. Topic theme-based sentiment summary.

We further investigate the weaknesses in terms of negative opinions by automatically mining the most frequent bigrams (i.e., a series of two adjacent words) in the review sentences. We separate those sentences in the corpus that contain patients' negative opinions in terms of their dissatisfaction with medical services. The preprocessed sentences are fed into the WEKA (Waikato environment for knowledge analysis) data mining toolkit to automatically mine the most commonly occurring bigrams. These bigrams point out the core problems related to that topic. After doing manual scrutiny, we considered only those bigrams that occur at least 20% in the review sentences. Only those bigrams which appear at least 20% in the review sentences were explored after the manual inspection. For instance, the most frequent bigrams occur for the weakness include, {('go,' 'hospital'), ('go,' 'early'), ('go,' 'with'), ('global,' 'crisis'), ('health,' 'crisis'), ('disease,' 'spread'), ('spread,' 'fast'), ('stop,' 'spread'), ('virus,' 'spread'), ('spread,' 'slow'), ('new,' 'cases'), ('report,' 'cases'), ('confirmed,' 'cases'), ('health,' 'emergency'), ('health,' 'services'), ('health,' 'officials'), ('health,' 'authorities'), ('health,' 'minister'), ('health,' 'care'), ('health,' 'workers'), ('health,' 'outcome'), ('public,' 'health'), and ('fine,' 'health')}.

According to a similar approach, we measure the patient dissatisfaction, keeping in view the weakness. The root cause analysis of patient dissatisfaction (negative opinion) is presented using the Ishikawa cause and effect diagram called the Fish-bone diagram (Figure 11). The main reasons for the weaknesses are established using the TSS, as revealed in Figure 10. Simultaneously, the secondary causes are determined using the bigram analysis and manual scrutiny of the keywords in review sentences.

During the COVID-19 outbreak, the secondary causes for the COVID-19 news were fake news, rumors spreading, mental health, day and night, and worse scenarios. Regarding the virus spread, patients expressed their displeasure toward new cases, death toll, positive ratio, and healthcare facilities. As far as the daily life impact is a concerned, patients felt unhappy about jobs lost, hike in goods prices, and food safety inside the restaurants. Finally, patients complain about the asymptotic virus symptoms after a negative nucleic acid test, such as frequent cough, body pain, sore throat, and abdominal discomfort (see Figure 11).

Apart from strengths and weaknesses, we uncovered several opportunities for the UK's healthcare system. During the COVID-19 crisis, hospital officials can design some training programs for healthcare providers for personality development and improving their communication skills. Doctors can implement new tools to improve system efficiency and treatment skills. To reduce simultaneous patient flow in the hospital during the pandemic, hospital officials can improve the hospital environment by improving the online appointment system and maintaining social distancing during business hours. Doctors can reduce treatment costs by introducing an online consultation system for the timely and economical delivery of medical consultation services. Hospital managers can introduce some training programs to enhance physicians' technical skills and functional skills of staff. During the COVID-19 crisis, patients' concerns toward several aspects of medical care services, that is, COVID-19 news, virus spread, impact of COVID-19 on daily life, nucleic acid test results an increase in competitors' strengths, which threats the physicians' service outcomes. Thus, such a development would be a major threat as it could negatively influence the competitive advantage.

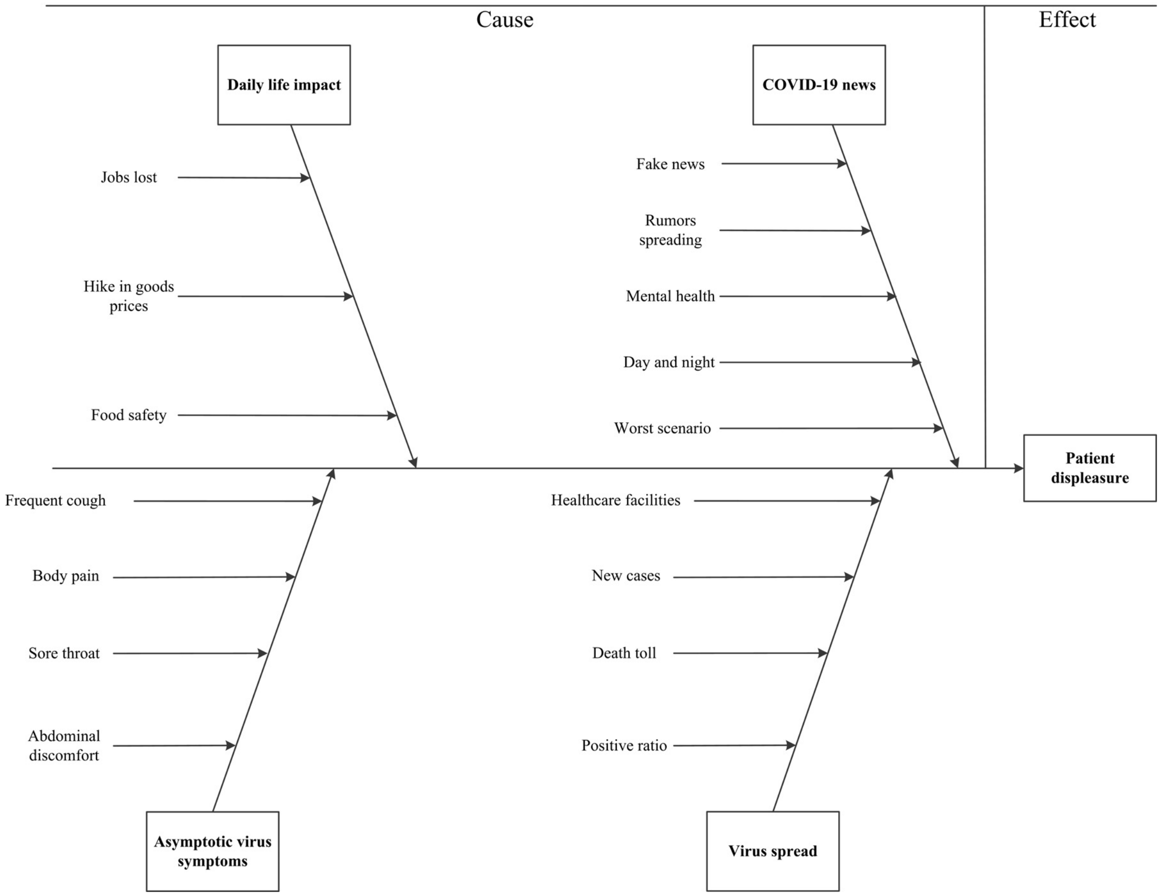

Figure 11. Fish-bone diagram of topics with overall negative sentiments.

CONCLUSION AND FUTURE WORK

This study reveals the systematic analysis of discussions and sentiments of PRW users to COVID-19 from February 01 to October 31, 2020, during three-time slices. Our results promote the swift and real-time interpretation of public conversations and feelings about the COVID-19 pandemic, supporting the monitoring system to figure out the evolving circumstances. The research overcomes the drawbacks of the conventional approach to social science, which focused on labor-intensive, observational, time-lagged, traditional surveys, and interviews. The trends and emotions defined by ODRs could be used to direct particular intervention programs. Our research contributes to the systematic analysis of ODRs using large-scale and real-time textual data from a UK based PRW in the following ways:

First, for early detection of the COVID-19 pandemic and a possible crisis in the UK, the initial suspected case was recorded in a woman (75 years old) in Nottinghamshire. This leads to an overwhelming number of ODRs posted on PRWs, which indicates that the health community recognized the seriousness of the disease even at the beginning of February.

Second, a small peak of the ODRs volume is identified between March 11 and March 15, and then gradually increases again after March 15, 2020. This result is also timed by the WHO declaration COVID-19 as a global pandemic. A growing volume of ODRs may be followed by the WHO declaration COVID-19 as a global pandemic, which gives a strong opportunity to direct the public to take precautionary action in March. The speed at which ODRs can be found posted on PRWs may lead the public and officials to react to the virus spread at the early phases.

Third, comments regarding the COVID-19 symptoms (breadth shortness, sore throat, and high fever) and treatments (e.g., vaccine, relax and sleep, drink plenty of water) were included in our collected ODRs during the first and second-time slice of the COVID-19, respectively. Findings from the previous study also indicated that for the COVID-19 symptoms (e.g., fatigue, pneumonia, illness, frequent cough), the volume of ODRs signals for symptoms rises over time (Abdellaoui *et al.*, 2018). These results also indicate that PRW is not commonly used as a forum for posting symptoms or seeking medical assistance or advice. Results suggest that more treatment-related reviews can be posted as an educational tool for the healthcare officials on PRW.

Fourth, our study identified new discussion topics about COVID-19 occurring from time slice 1 to time slice 3: (1) news about the COVID-19 pandemic, (2) quarantine process and quarantine centers at different locations, (3) violence against the lockdown, and (4) vaccine development/treatment to stop virus spread.

Fifth, the new salient topics suggest that PRW users in UK are focusing their attention on COVID-19 (e.g., news, quarantine policy, violence, vaccine trials) rather than global news (e.g., South Korea, Protest against the black life matter, Diamond Princess Cruise ship, and Dr. Li Wenliang in China). Due to existing preventive initiatives, hand washing is no longer a problem, but the economic effects and production of vaccines have instead become predominant during the third time slices of the COVID-19.

Sixth, our findings show that negative emotions are dominant during the COVID-19 epidemic, supporting the recent call for action to preserve the mental well-being of the public for this unprecedented mischievous crisis. Fear reveals as a leading emotion in all topics during all three-time slices of the COVID-19 crisis. These findings are in line with previous results (Lwin *et al.*, 2020; Xue *et al.*, 2020a; Su *et al.*, 2020; Xue *et al.*, 2020b), which indicates that COVID-19 has a significant effect on the psychological state of the public. Opinion mining using the COVID-19 data helps us to understand the dynamics of public concerns and sentiments during the crisis. Our results have implications for health departments, essential for mental health and psychosocial well-being assistance required during the current pandemic crisis (Griffis *et al.*, 2020).

Finally, we conducted a SWOT analysis using TSS. The strengths and weaknesses of the doctors and the healthcare system were respectively based on patients' appreciation and displeasure. In contrast, the opportunities and threats were identified by examining topics related to patient dissatisfaction. Our study findings indicate that ODRs represent an emerging online asset that closely resembles the real healthcare system. As a result, ODRs could help healthcare providers better understand patients' voices in terms of delight and anger to improve the quality of care. On the other hand, patients can take advantage of this publicly available online asset to choose a competent doctor that would provide the best medical services for their disease treatment. Overall, ODRs are a win-win asset for both the doctors and their patients in the UK.

Despite the number of contributions, there might be several limitations that may lead to the current study's possible extensions. First, the demographics of PRWs users may not necessarily represent the populations of the selected cities in the UK, which limits the generalizability of our findings. Even though this website is one of the popular websites in the UK Future research can include more reviews and doctors from other regions to improve the generalizability of the study. Second, future research should further examine sentiments by investigating the other media outlets from particular countries. Third, our choice to include only English ODRs means that non-English speakers have not been represented.

Nevertheless, our techniques are consistent with clinical research that makes use of topic modeling of ODRs and is mostly confined to the country's official language (Griffis *et al.*, 2020). Finally, the patient perceived SWOT analysis can also be compared across different demographics: patients' age, gender, health status, and level of the hospital. For example, which patient demographic group or level of the hospital is influenced by positive or negative online word of mouth? In summary, our findings indicated that analysis of online textual reviews using the automated hybrid approach could provide exciting insights for patients regarding their satisfaction and dissatisfaction toward the quality of care and for practitioners to take several steps in improving their healthcare system using advanced computational techniques.

ACKNOWLEDGMENT

This work is partly supported by the National Natural Science Foundation, of People's Republic of China (No. 72025101, 71729001).

REFERENCES

Abdellaoui, R., Foulquié, P., Texier, N., Faviez, C., Burgun, A., & Schück, S. 2018. Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts: Topic Model Approach. J Med Internet Res, **20**(3): e85.

Ali, K.F., Whitebridge, S., Jamal, M.H., Alsafy, M., & Atkin, S.L. 2020. Perceptions, Knowledge, and Behaviors Related to COVID-19 Among Social Media Users: Cross-Sectional Study. J Med Internet Res, **22**(9): e19913.

Bird, S., Klein, E., & Loper, E. 2009. Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit. "O'Reilly Media, Inc.", Massachusetts, United States.

Blei, D.M., & Lafferty, J.D. 2006. Dynamic topic models, In Proceedings of the 23rd international conference on Machine learning, pp. 113–120. ACM. Pittsburgh, Pennsylvania, USA.

Blei, D.M., Ng, A.Y., & Jordan, M.I. 2003. Latent Dirichlet Allocation. J Mach Learn Res, **3**: 993-1022.

Boyd-Graber, J., Hu, Y., & Mimno, D. 2017. Applications of Topic Models. Foundations and Trends® in Information Retrieval, **11**(2-3): 143-296.

Cao, L., & Tang, X. 2014. Topics and trends of the on-line public concerns based on Tianya forum. Journal of Systems Science and Systems Engineering, **23**(2): 212-230.

Chen, E., Lerman, K., & Ferrara, E. 2020a. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health Surveill, **6**(2): e19273.

Chen, L., Lyu, H., Yang, T., Wang, Y., & Luo, J. 2020b. In the Eyes of the Beholder: Analyzing Social Media Use of Neutral and Controversial Terms for COVID-19. arXiv preprint arXiv:200410225.

Colnerič, N., & Demšar, J. 2020. Emotion Recognition on Twitter: Comparative Study and Training a Unison Model. IEEE Transactions on Affective Computing, **11**(3): 433-446.

Convertino, J. 2020. ABC News, Social media companies partnering with health authorities to combat misinformation on coronavirus. https://abcnews.go.com/Technology/social-media-companies-partnering-health-authorities-combat-misinformation/

story?id=69389222.

Dong, M., Cao, X., Liang, M., Li, L., Liang, H., & Liu, G. 2020. Understand Research Hotspots Surrounding COVID-19 and Other Coronavirus Infections Using Topic Modeling. medRxiv.

Doogan, C., Buntine, W., Linger, H., & Brunt, S. 2020. Public Perceptions and Attitudes Toward COVID-19 Nonpharmaceutical Interventions Across Six Countries: A Topic Modeling Analysis of Twitter Data. J Med Internet Res, **22**(9): e21419.

Gensim. **2019.** Dynamic Topic Models (DTM) and Dynamic Influence Models (DIM). https://radimrehurek.com/gensim/models/wrappers/dtmmodel.html.

Gerrish, S., & Blei, D.M. 2010. A Language-based Approach to Measuring Scholarly Impact, In Proceedings of the 27th International Conference on Machine Learning, pp. 375-382. Omnipress. Haifa, Israel.

Gozzi, N., Tizzani, M., Starnini, M., Ciulla, F., Paolotti, D., Panisson, A., & Perra, N. 2020. Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis. J Med Internet Res, **22**(10): e21597.

Griffis, H., Asch, D.A., Schwartz, H.A., Ungar, L., Buttenheim, A.M., Barg, F.K., Mitra, N., & Merchant, R.M. 2020. Using Social Media to Track Geographic Variability in Language About Diabetes: Infodemiology Analysis. JMIR Diabetes, **5**(1): e14431.

Grimmer, J., & Stewart, B.M. 2017. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Polit Anal, **21**(3): 267-297.

Gui, X., Wang, Y., Kou, Y., Reynolds, T.L., Chen, Y., Mei, Q., & Zheng, K. 2018. Understanding the Patterns of Health Information Dissemination on Social Media during the Zika Outbreak. AMIA Annual Symposium proceedings AMIA Symposium, **2017**: 820-829.

Gupta, R.K., & Yang, Y. 2018. CrystalFeel at SemEval-2018 Task 1: Understanding and Detecting Emotion Intensity using Affective Lexicons, In Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval-2018), pp. 256-263. ACL. New Orleans, Louisiana.

Ha, T., Beijnon, B., Kim, S., Lee, S., & Kim, J.H. 2017. Examining user perceptions of smartwatch through dynamic topic modeling. Telematics and Informatics, **34**(7): 1262-1273.

Househ, M. 2016. Communicating Ebola through social media and electronic news media outlets: A cross-sectional study. Health Informatics Journal, **22**(3): 470-478.

Hung, M., Lauren, E., Hon, E.S., Birmingham, W.C., Xu, J., Su, S., Hon, S.D., Park, J., Dang, P., & Lipsky, M.S. 2020. Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence. J Med Internet Res, **22**(8): e22590.

Jang, H., Rempel, E., Carenini, G., & Janjua, N. 2020. Exploratory Analysis of COVID-19 Related Tweets in North America to Inform Public Health Institutes. arXiv preprint arXiv:200702452.

Janz, N.K., & Becker, M.H. 1984. The Health Belief Model: A Decade Later. Health Education Quarterly, **11**(1): 1-47.

Jelodar, H., Wang, Y., Orji, R., & Huang, H. 2020. Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. arXiv preprint arXiv:200411695.

Jiaxiang, L., Shuyi, W., & Svitlana, G. 2020. Dynamic Topic Modeling. https://pypi.org/project/dynamic-topic-modeling/.

Kim, T.J. 2020. COVID-19 News Analysis Using News Big Data: Focusing on Topic Modeling Analysis. The Journal of the Korea Contents Association, **20**(5): 457-466.

Li, J., Xu, Q., Cuomo, R., Purushothaman, V., & Mackey, T. 2020. Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study. JMIR Public Health Surveill, **6**(2): e18700.

Luke, L. 2018. Exploring the UN General Debates with Dynamic Topic Models. https://towardsdatascience.com/exploring-the-un-general-debates-with-dynamic-topic-models-72dc0e307696.

Lwin, M.O., Lu, J., Sheldenkar, A., Schulz, P.J., Shin, W., Gupta, R., & Yang, Y. 2020. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. JMIR Public Health Surveill, **6**(2): e19447.

Mihalcea, R., & Tarau, P. 2004. Textrank: Bringing Order Into Texts, In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp. 404-411. ACL. Barcelona, Spain.

Ordun, C., Purushotham, S., & Raff, E. 2020a. Exploratory Analysis of Covid-19 Tweets using Topic Modeling, UMAP, and DiGraphs. arXiv preprint arXiv:200503082 https://arxiv.org/pdf/2005.03082.pdf.

Ordun, C., Purushotham, S., & Raff, E. 2020b. Exploratory Analysis of Covid-19 Tweets using Topic Modeling, UMAP, and DiGraphs, In arXiv preprint arXiv:2005.03082.

Parra, D., Trattner, C., Gómez, D., Hurtado, M., Wen, X., & Lin, Y.-R. 2016. Twitter in academic events: A study of temporal usage, communication, sentimental and topical patterns in 16 Computer Science conferences. Computer Communications, **73**: 301-314.

Paul, M.J., & Dredze, M. 2011. You Are What You Tweet: Analyzing Twitter for Public Health, In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, pp. 1-8. AAAI. Barcelona, Spain.

Paul, M.J., & Dredze, M. 2014. Discovering health topics in social media using topic models. PloS One, **9**(8): e103408.

Plutchik, R. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American Scientist, **89**(4): 344-350.

Porter, M.F. 2001. Snowball: A Language For Stemming Algorithms.

Rehurek, R. 2014. Gensim.models.coherencemodel – Topic coherence pipeline. https://radimrehurek.com/gensim/models/coherencemodel.html.

Röder, M., Both, A., & Hinneburg, A. 2015. Exploring the space of topic coherence measures, In Proceedings of the eighth ACM international conference on Web search and data mining, pp. 399-408. ACM. Shanghai, China.

Saleh, S.N., Lehmann, C.U., McDonald, S.A., Basit, M.A., & Medford, R.J. 2020. Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing on Twitter. Infection Control and Hospital Epidemiology https://doi.org/10.1017/ice.2020.406: 1-8.

Sha, H., Hasan, M.A., Mohler, G., & Brantingham, P.J. 2020. Dynamic topic modeling of the COVID-19 Twitter narrative among US governors and cabinet executives. arXiv preprint arXiv:200411692.

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. 2012. Exploring topic coherence over many models and many topics, In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 952-961. ACL. Jeju Island, Korea.

Stokes, D.C., Andy, A., Guntuku, S.C., Ungar, L.H., & Merchant, R.M. 2020. Public Priorities and Concerns Regarding COVID-19 in an Online Discussion Forum: Longitudinal Topic Modeling. J Gen Intern Med, **35**(7): 2244-2247.

Su, Y., Xue, J., Liu, X., Wu, P., Chen, J., Chen, C., Liu, T., Gong, W., & Zhu, T. 2020. Examining the Impact of COVID-19 Lockdown in Wuhan and Lombardy: A Psycholinguistic Analysis on Weibo and Twitter. International Journal of Environmental Research and Public Health, **17**(12): 4552.

WHO. 2020a. Coronavirus Disease (COVID-19) Pandemic. https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

WHO. 2020b. Coronavirus Disease (COVID-2019) Advice for Public. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public.

WHO. 2020c. WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/.

WHO. 2020d. WHO Coronavirus Disease (COVID-19) Dashboard for Great Britian. https://covid19.who.int/region/euro/country/gb.

Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. 2020a. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PloS One, **15**(9): e0239441.

Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., & Zhu, T. 2020b. Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. J Med Internet Res, **22**(11): e20550.

Yin, H., Yang, S., & Li, J. 2020. Detecting Topic and Sentiment Dynamics Due to COVID-19 Pandemic Using Social Media. arXiv preprint arXiv:200702304.

Zhao, Y., Cheng, S., Yu, X., & Xu, H. 2020. Chinese Public's Attention to the COVID-19 Epidemic on Social Media: Observational Descriptive Study. J Med Internet Res, **22**(5): e18825.

Zhu, B., Zheng, X., Liu, H., Li, J., & Wang, P. 2020. Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics. Chaos, Solitons & Fractals, **140**: 110123.