Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Analyzing the attitude of Indian citizens during the second wave of COVID-19: A text analytics study

Surjit Paul

Vinod Gupta School of Management, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, 721302, India

ARTICLE INFO

Keywords:
COVID-19
Second wave
Text analytics
Sentiment analysis
Topic modeling

ABSTRACT

Background and aims: The COVID-19 pandemic outbreak has created severe public health crises and economic consequences across the globe. This study used text analytics techniques to investigate the key concerns of Indian citizens raised in social media during the second wave of COVID-19.

Methods: In this study, we performed a sentiment and emotion analysis of tweets to understand the attitude of Indian citizens during the second wave of COVID-19. Moreover, we performed topic modeling to understand the significant issues and concerns related to COVID-19.

Results: Our results show that most social media posts were in neutral tone, and the percentage of posts that showed positive sentiment was less. Furthermore, emotion analysis results show that ‘Fear’ and ‘Surprise’ were the prominent emotions expressed by the citizens. Topic modeling results reveal that ‘High crowd’ and ‘political rally’ are the two primary topics of concern raised by Indian citizens during the second wave of COVID-19.

Conclusions: Hence, Indian government agencies should communicate crisis information and combating strategies to citizens more effectively in order to minimize the fear and anxiety amongst the public.

1. Introduction

The COVID-19 pandemic outbreak has resulted not only in severe public health crises but has created substantial economic consequences across the globe [28]. Due to the capacity to produce severe health consequences and its highly contagious nature, COVID-19 is termed a once in a century medical calamity [33]. No such event has ever happened in the past hundred years of human civilization that has created such havoc as COVID-19 [34]. Many people from different countries got infected and succumbed to this deadly virus. India became the second most affected country after the United States as on April 12, 2021. The number of COVID-19 new cases and death cases started rising in early April 2021 and peaked on May 6, 2021. The COVID-19’s wave created havoc in terms of total cases and fatalities across the globe. The total number of COVID-19 confirmed cases in India was 30,411,634, and the death toll climbed to 399,459 as of June 30, 2021, whereas the total cases and fatalities were merely 585,481 and 17,400, respectively, in the previous year. However, the fatality rate was low, close to 1.2%, which has been hypothesized by many Indian researchers and later validated [38] but never highlighted in News channels. Similar to other countries, India faced a severe health crisis during the second wave of COVID-19, which put immense pressure on the existing healthcare system. Hospitals in India were overwhelmed and...
struggling to cope with critical drugs and oxygen supply. The government of India directed different state governments to impose night curfews or lockdowns to minimize the surge of COVID-19 new cases.

Many countries such as the United States, Brazil, the UK, and Germany faced the second wave in early November 2020. Although these countries encountered healthcare challenges, the Indian healthcare system was not even prepared enough to tackle the worst impact of the second wave. Citizens of India had shown evident distress against the existing healthcare system of India during the second wave. Hence, it is crucial to understand peoples’ attitude, behavior, and opinions during such a crisis.

This project is partially funded by University Grants Commission under NET-JRF scheme (Grant No: F.15–9 (June 2015))/2015 (NET).

Nowadays social media has become a prominent medium to disseminate information [17,19]. As a consequence of the COVID-19 pandemic, social media has emerged as a pivotal platform of choice for public opinions, views, and attitudes towards various COVID-19 related guidelines or public health policies [35].

Despite the low case fatality rate, which is lower than many other health illnesses relevant to the Indian scenario (such as tuberculosis which carries a fatality rate of >5% in India and 7–35%. Globally [39]), the Twitter was flooded with posts. Emerging contagious diseases, such as COVID-19, resulted in a huge surge in usage and consumption of social media by the common public for information [36]. A plethora of studies in the literature that have analyzed social media posts on various social platforms such as Twitter, Facebook, and blogs to understand the attitude of people on different crises [9,19,23]. A study has also surveyed amongst the respondents of India, Japan, Philippines, and the US. They found that respondents from these countries commonly stated that COVID-19 impacted “mental health/experienced stress and trauma” and “reduced social ties” [37].

Given the severity and consequences of the COVID-19 pandemic in the Indian context, a detailed study is needed to gain insights into citizens’ attitudes and behavior in social media. However, limited studies have attempted to understand citizens’ attitudes, behavior, and opinion in social media during the second wave of the COVID-19 pandemic in the Indian context. Since the threat of such future pandemics cannot be fully ruled out, it is important to understand the Indian citizens’ viewpoints in social media during the second wave of COVID-19. Therefore, this study attempts to analyze Indian citizens’ attitudes and perspectives during the second wave of COVID-19.

Furthermore, we try to find key concerns of Indian citizens’ voiced out in social media during the second wave. This study has implications for government officials and policymakers to understand common citizens’ concerns to combat such future pandemics. Therefore, the current research tries to answer the following research questions:

RQ 1. What kind of attitude and behavior did Indian citizens display during the second wave of COVID-19?

RQ 2. What were the key concerns expressed by Indian citizens during the second wave of COVID-19?

2. Methodology and data

2.1. Sentimental analysis

Twitter sentiment analysis has gained significant popularity amongst information system researchers due to its applications in public opinion monitoring and disaster warnings [16,24,32]. Many prior studies have conducted sentiment analysis on Twitter data [2,20,32]. Hence, to address our first research question, we performed sentiment analysis on our tweets dataset to understand the attitude and behavior of Indian citizens during the second wave of COVID-19. Sentiment analysis is used to compute the polarity of text content to identify whether the sentiment of the content is positive, negative, or neutral. It is a machine learning technique that uses NLP (Natural Language Processing) to identify the sentiments expressed by authors in a particular text [7,18]. Sentiment analysis is described as “the automatic method to extract and analyze the subjective judgments on different aspects of an item or entity” [25]. Previous literature highlighted three different levels of sentiment analysis, such as document-level [21,26], sentence-level [15], and phrase-level [1,29,30]. Sentiment analysis of tweets are usually conducted using rule based sentiment lexicons or machine learning methods, such as Maximum Entropy, Naive Bayes, and Support Vector Machine (SVM) algorithms. However, this study applied VADER (Valence Aware Dictionary and sEntiment Reasoner) python library [13], a rule based sentiment analysis having 96% classification accuracy, for analyzing our tweets dataset to measure the overall sentiment score conveyed. It was found that VADER lexicon performed exceptionally well compared to other machine learning methods in the social media domain [13].

Since sentiment analysis merely report polarity of text content whether positive, negative, or neutral, few prior studies analyzed emotions expressed by the authors in the text content [3,11]. Hence, this study also examined the emotions expressed in each tweet using text2emotion Python Library. This library computes five different emotion categories of the text content such as ‘Happy’, ‘Angry’, ‘Surprise’, ‘Sad’, and ‘Fear’.

2.2. Topic modeling

Since sentiment analysis does not explore the concerning topics or concerns that shape the attitude of Indian citizens during the second wave of COVID-19, to address the second research question we used Latent Dirichlet Allocation (LDA) topic modeling method [22]. LDA helps to discover latent topics from a large unstructured data. It was first introduced by Blei, Ng, & Jordan (2003). LDA topic modeling is a technique that explores topics and themes hidden inside a set of corpora using a set of algorithms [3]. However, prior to the introduction of LDA, Latent Semantic Indexing (LSI) was proposed [10] which uses a singular value decomposition of the X matrix to recognize a linear subspace in the space of tf-idf features that capture most of the variance.

Further, Probabilistic Latent Semantic Indexing (pLSI) method was introduced [14]. The basic theory behind pLSI is that it models each word in a document as a sample from a mixture model and the mixture components are the multinomial random variables, which
can be represented as topics. However, pLSI does not provide a probabilistic model at the level of documents [4].

Nonetheless, LDA follows the “Bag of words” assumption, which represents a document as a mixture of latent topics in which a topic is a probabilistic mixture of words. The primary aspect of LDA topic modeling is that the algorithms assume that all the documents in a corpus display a similar set of topics and each document shows a various probabilistic mixture of those topics.

2.3. Data collection

After the outbreak of COVID-19, citizens’ across the globe started posting their concerns about the COVID-19 pandemic on Twitter [31]. Based on prior studies, we can infer that social media is the best source from which we can understand the general public’s attitude and behavior during crisis time [6,8,12,27]. Hence, in order to understand common mental attitudes of Indian citizens during the second wave of COVID-19 pandemic, we scrapped tweets posted on Twitter between March 25, 2021 and June 30, 2021 containing the words “Covid19” using the Python Twint library. We chose this timeframe because of two reasons. First, during that period, the number of COVID-19 daily cases, hospitalization, and fatalities in India increased significantly compared to the previous year. Second, the number of tweets posted related to “Covid19” also increased during that period.

Since the Twint library provides an option of geographical filtering, we only gathered tweets posted by Indian citizens. Further, we considered tweets that were written in the English language. Finally, after filtering the tweets based on both criteria, the final tweets dataset contained 30,168 unique tweets for the study. Then we performed data preprocessing and cleaning to remove stop words such as numbers, punctuations, and hyperlinks from the corpus that were not needed for our data analysis. Further, we conducted stemming and lemmatizing on our dataset. Stemming eliminates terminations of terms to identify their root form, while lemmatization considers the context and converts the word to its meaningful base form.

3. Results

3.1. Sentimental analysis

Fig. 1 shows the COVID-19 daily new cases during the second wave of the COVID-19 pandemic. The maximum number of daily cases of COVID-19 reported in India was 414,188 as of May 6, 2021. Fig. 2 shows the month-wise number of tweets posted related to COVID-19. The majority of tweets were posted in April and May 2021. Fig. 3 shows the graphical representation of the sentiment analysis results. It shows that neutral emotion was dominant in our tweets dataset, followed by positive and negative sentiment. We observed that positive emotion got reduced by approximately 2% in May compared to previous months. Also, there was an increase in negative tone by 2.4% in the month of April compared to March 2021. Hence, we can infer a negative correlation between the positive sentiment of tweets and number of COVID-19 daily cases. Further, the trend also showed a positive correlation between number of cases and negative emotions in the Months of March–April 2021.

Further, emotion analysis results showed that ‘Fear’ (45.6%) and ‘Surprise’ (27.2%) were the prominent emotions followed by ‘Sad’ (10.3%) expressed by majority of tweets during the second wave of COVID-19.

3.2. Topic modeling

Though sentiment analysis provided an insight into people’s attitudes during the second wave of COVID-19, it did not give any information about the latent themes inside an unstructured text content. Hence, to explore the hidden topics inside our tweets dataset, we further performed LDA topic modeling. We explored ten prominent topics from our tweets dataset. Table 1 displays the results of the LDA topic modeling.
Discussion

Although most tweets regarding the second wave of COVID-19 exhibit neutral sentiment, the percentage of tweets showing positive and negative emotion is less. Hence, we can conclude that most people in India did not post negative tweets on Twitter. Though tweets showing negative emotion are still there, it provides a direction to the government to improve crisis mitigation strategies. Moreover, tweets showing positive sentiment are minimal, and ‘Fear’ and ‘Surprise’ are the prominent emotions expressed by each tweet. Hence, findings suggest that government agencies should take appropriate strategies to communicate their crisis information and mitigation action plans to minimize the negative sentiment and increase the positive sentiment. These plans would eventually increase the trust of the people in government agencies and reduce fear, distress, and anxiety during an emergency.

Further, topic modelling results reveal ten prominent topics: board exam cancellation, Mumbai COVID-19 surge mitigation strategy, political rallies, Pune city COVID-19 update, COVID-19 statistics, and COVID-19 vaccine awareness, high crowd, COVID-19 guidelines, Chennai COVID-19 update, Delhi school opening. Table 1 presents the topic modeling results.

| Topic                                         | Words                                                                                     |
|-----------------------------------------------|-------------------------------------------------------------------------------------------|
| Board exam cancellation                       | lockdown, cancelboard, drprnishank, manage, student, wearamask                             |
| Mumbai COVID-19 surge mitigation strategy     | mumbai, minister, invest, inform, night, surge, maskup, rule, mybmc                         |
| Political rallies                             | maharashtra, karnataka, time, country, rally, look, officeofut, contain                    |
| Pune city COVID-19 update                     | total, toll, pune, city, update, place, person                                            |
| COVID-19 statistics                           | case, death, report, test, recovery, active, people, state                                |
| COVID-19 vaccine awareness                    | vaccine, india, coronavirus, today, narendramodi, corona, take, stock, health             |
| High crowd                                    | opinion, political, get, request, allow, single, high, crowd                              |
| COVID-19 guidelines                           | mask, wear, elect, spread, market, write, public, issue                                    |
| Chennai COVID-19 update                       | posit, covidecondwave, increase, chennai, staysafe, good, control                         |
| Delhi school opening                          | delhi, technology, start, school, care, wait, avoid, arvindkejriwal                         |

Fig. 2. Month-wise posted tweets related to COVID-19.

Fig. 3. Graphical representation of month-wise sentiment analysis.

Table 1

LDA topic modeling results.
guidelines, Chennai COVID-19 update, and Delhi school opening. Our findings highlight themes related to COVID-19 statistics, vaccination, students’ issues, city-specific COVID-19 updates and mitigation strategies, and public concerns of COVID-19 surge due to political rallies during the second wave of COVID-19. Our study could assist government agencies to understand the issues and concerns raised by the people in social media and devise appropriate action plans to minimize the public’s fear, distress, and anxiety during a crisis.

4.1. Practical implications

The implications of our study are twofold. First, we have performed sentiment analysis and emotion analysis of tweets posted by Indian citizens related to COVID-19 during the second wave of COVID-19. We computed month-wise sentiment of posted tweets to understand the attitude of Indian citizens. Our findings reveal that a negative correlation between the positive sentiment of tweets and an increase in the number of COVID-19 daily cases during our sample period. Furthermore, ‘Fear’ and ‘Surprise’ are the prominent emotions expressed by the majority of tweets, followed by ‘Sad’. One of the possible reasons for fear is the continuous bombarding of news related to COVID-19 fatalities by news channels, without highlighting the low case fatality rate of the disease, which was rampant throughout the pandemic in India [38]. Hence, government agencies should communicate crisis information and combating strategies more effectively to the public to increase the trust of people in government. Second, since the threat of such future pandemics could not be entirely ruled out, it is crucial to understand the Indian citizens’ perspectives to COVID-19’s wave.

We explored ten prominent themes based on our tweets dataset. Our study could assist Indian government agencies in understanding the public’s issues and concerns raised in social media during the second wave of COVID-19 and devise proactive measures to minimize the fear and anxiety amongst the public and delay such future pandemics.

4.2. Limitations and future research

This study has certain limitations. First, the present study focused only on Twitter posts to analyze the attitude and perspective of Indian citizens during the second wave of COVID-19. However, other social media platforms can be utilized for comprehensive analysis. Therefore, in future studies, we could try to analyze social media posts on various social media platforms to understand the effectiveness of different media platforms for the diffusion of information related to the crisis. Second, this study performs sentiment analysis of posted tweets related to COVID-19 between March 25, 2021 and June 30, 2021 in the Indian context. Hence, it cannot be extrapolated to other populations. Further, the analysis focused on the second wave of pandemic, so could not generate information specific to first or third wave. Hence, future studies might explore the sentiment and latent themes present in the social media content for different crises and country contexts. Finally, the study does not include opinions of subgroup experts (based on Twitter users’ profiles), legal experts, human rights activists, and medical personnel that would provide more focused information about concerns in respective fields and would also be relevant in future decision plans.

5. Conclusion

Social media plays a crucial role in rapid information diffusion during an emergency. This study uses sentiment & emotion analysis, and topic modeling methods to understand the attitude and perspectives of people during the second wave of COVID-19 in the Indian context. Our findings reveal that there is a negative correlation between the positive sentiment of tweets and an increase in the number of COVID-19 daily cases. Emotion analysis results highlight that fear and surprise were the predominant emotions. Hence, government agencies should communicate crisis information and combating strategies more effectively to the public to increase the trust of the people in government. Furthermore, we find ten prominent themes using topic modeling. Our study could assist Indian government agencies in understanding the public’s issues and concerns raised in social media during the second wave of COVID-19 and devise proactive measures to minimize the fear and anxiety amongst the public and delay such future pandemics. However, future studies need to find the source of increased fear in public, highlight the low case fatality rate of COVID-19, and appropriate strategies on the part of government agencies to regulate the continuous negative broadcast by news channels.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] A. Agarwal, F. Biadsy, K. Mckeown, Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams, in: Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), 2009, pp. 24–32.
[2] A. Bermingham, A.F. Smeaton, Classifying sentiment in microblogs: is brevity an advantage?, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, 2010, pp. 1833–1836.
[3] S. Bhooshan, R. Praveen Pai, R. Nandakumar, A sentiment analysis of a boycott movement on twitter, in: Computer Networks and Inventive Communication Technologies, Springer, Singapore, 2022, pp. 313–324.
[4] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[5] D.M. Blei, J.D. Lafferty, Topic models, in: Text Mining, Chapman and Hall/CRC, 2009, pp. 101–124.
