Twitter Sentiment Analysis During Unlock Period of COVID-19

Swati Sharma
Department of Humanities and Social Science
Jaypee Institute of Information Technology Noida
Noida, India
sharma.a.swati@gmail.com

Aryaman Sharma
Department of Computer Science
Netaji Subhash Institute ofTechnology
Delhi, India
aryamansharma01@gmail.com

***Abstract*— The pandemic has hit the individuals at both personal, social and professional front triggering emotional crisis leading to stress, anxiety and other related problems. However, some countries are now easing down on restrictions by going from lock down to unlocking in a phased manner. As life springs back to action the sentiments and emotions of people are bound to change. It therefore becomes imperative to understand the emotions and sentiments of people after seven months of outbreak when the people are more informed about the nature of disease, steps for prevention and also have hope for a vaccine coming up in near future. The study analyses the sentiments of the people from the USA and India by text mining using R Studio. The study has various implications for academicians as it adds to the existing knowledge pool. The findings provide guidance to the policy makers to tailor their support policies in response to the emotional state of their people and also assists the marketers to tailor the communication strategies in the light of the emotional state of the target market.**

Keywords*— *COVID-19, Sentiment Analysis, Twitter, Rstudio, USA, India

I. INTRODUCTION

COVID-19 outbreak was declared a Public International Concern by the WHO on January 30, 2020, spreading in almost every country in the world, reaching pandemic proportions. This outbreak of virus, follow by government imposed restrictions on movement (lockdown), social distancing and rising death cases threatened not only people's physical health, but also affected their mental health, especially in terms of emotions [1]. This pandemic resulted in an incomparable impact on global economy and human health both physical and emotional [2]. Studies conducted on emotions with the outbreak of COVID cited a rise in negative emotions and pessimism among people [1]. However, after 7 months of outbreak, the rising corona cases are also accompanied with greater information on how it spreads and how it can be prevented. Also the news of vaccines under development, with 26 of these in the human trial phase (www.who.int) create a certain level of confidence among the people. The events of phased opening of market places , beaches , schools in some countries, etc are signals of people embracing the new normal by developing resilience.

The paper studies the related work on sentiments during COVID and presents the sentiment analysis using twitter data after 7 months of pandemic outbreak throwing light on community resilience.

II. RELATED PAST WORK ON EMOTIONS AND COVID-19

Emotions have been central to the COVID-19 related research among others. Researchers in the context of COVID-19 have extensively studied the rise in negative emotions [3],[4] and their regulation [2] . Researchers have reinforced emotional and behavioural changes among people as a response to pandemic and also among others focused on resilience [5]. Resilience is the ability of a society or geographically defined area to deal with crisis and efficiently resume daily life [6].

Research provides that most people are resilient when facing highly aversive life. Disaster research has acknowledged that communities regularly work together to survive and recover from tragic impacts [7]. The concept of community resilience has been drawn by researchers of diverse fields such as psychology [8],[9], sociology [10], socioecological systems [11], natural disaster research[12] and health crisis [13],[14],[15],[16]. Research provides that most people are resilient when facing a highly aversive life. Several nationwide cross-sectional studies have documented depression and anxiety symptoms among people during the COVID-19 pandemic [17],[18] and how adopting precautionary measures respondents experienced less psychological stress. Emotions and sentiments affect stress coping abilities. Positive emotions enhance psychological well-being and negative emotions adversely affect the wellbeing and stress coping abilities [19],[20],[5]. Researchers have extensively studied people's emotions and sentiments during the present pandemic crisis. A recent study probed emotions before and after the announcement of COVID [1]. Their findings revealed that pandemic showed the rise in negative emotions among the weibo users. This study concentrated on Chinese weibo users only. The present study addresses this limitation and uses big data to analyze the emotions and sentiments of the people using twitter data extracted from various locations and sources. The study builds on the premise that with general tendency of people to bounce back in the face of the adversity [6] creates a need to study the emotional patterns of the society.

With seven months on struggling to adjust to the "new normal" of pandemic life and the tendency to resile [6] in crisis situations this study aims to analyze the emotions and sentiments of people after their coping the pandemic. It is essential to study the present sentiments of the people as nations plan to revive economies and strategize the revival. The objective of this paper is to analyze the sentiments of the people after coping with pandemic for 7 months with special focus on India and USA. The two countries leading in the number of reported cases.

III. RESEARCH METHODOLOGY

This study uses a popular microblog called Twitter to study the sentiments of the bloggers regarding COVID -19 during the unlock period. Our methodology consists of the following steps:

2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC) | 978-1-7281-7132-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/PDGC50313.2020.9315773

A. *Extraction of tweets*

The template is used to format your paper and style the text. All a twitter application was created. Twitter packages act as an interface to the Twitter API. rtweet package is used for authentication During authentication, redirection to a URL is done automatically by clicking on the authorized app, and verified using captcha. Raw data was extracted by using function search tweets using different queries such as covid-19, COVID, corona, corona virus, pandemic, and an intersection of tweets contain Wuhan and virus. The raw data consisted of the union of the tweets extracted using these queries. 83716 observations were extract.

B. *Data Pre-processing*

The data was processed by selecting specific attributes such as status id, time, screen named, text, source and location. The resultant data was further filtered by removing unnecessary data in the text such as HTML links, emoticons, punctuations '@', stopwords (i.e. is, at the, on), RT, numbers, white spaces were removed.

C. *Analyses*

A plot was constructed to identify the top 20 locations from where the tweet was posted. The plot revealed the origin of tweets where the majority of the tweets were from USA (1314 tweets) and India (1212 tweets) the two big countries with highest COVID-19 cases. The remaining tweets were scattered or did not have a location attached to it. The data was categorized into two labels i.e. United States and India The words from all the tweets were extracted separately for each category and their corresponding frequency was tabulated and a wordcloud was created. The pre-processed tweets were scored and classified by polarity (positive or negative) and categorized into 10 different types of emotions using the R package titled "Gutenberger" and NRC emotion-based dictionary. Sentiment analysis was done by identifying the polarity in the rage of $-5 \leq v \leq 5$ (AFINN 2011). Emotions of words in tweets were analyzed as expressed in overall corpus review data by using NRC Word-Emotion Association Lexicon (NRCLex Documentation, 2020) as developed by Mohammad, & Turney [21]. Sentiment extracted (positive, or negative) from AFINN analysis was categorized and a wordcloud for the same were generated for each category. There were 45,256 positive and 71,644 negative sentiments.

IV. FINDINGS AND DISCUSSIONS

After pre- processing of the raw data the total tweets studied were 83716. The tweets selected for the study were further segregated on the basis of source or location. The two major countries which are leading in the number of reported cases were selected with 1314 tweets from USA and 1212 tweets from India. The number of words extracted from the tweets were tabulated on the basis of their frequency. It was observed that the words "COVID", "CONORA Virus" "Pandemic" and "CORONA" had exceptionally high frequency. This was because the tweets were extracted using these keywords. Since this is a known context of the study and these exceptionally high frequency words were eliminated and treated as outliers for making of the wordcloud. The word cloud was prepared excluding them so that the meaningful information could be extracted. Wordcloud of positive and negative words from the total corpus categorised location wise is presented. Figure1 and 2 represents the wordcloud for positive and negative words for India and figure 3 and 4 represent for USA data. A wordcloud of total words (both positive and negative) for India and US is as given in figure 5 and figure 6 respectively.

According to the first post.com The United States is the world's worst-affected country, with corona cases exceeding 6.3 million and India is second in terms of total cases recorded. To fulfil the research objective i.e. to understand the emotions of two most affected countries, the data was segregated region wise and the tweets with origin of US and those of India were analyzed. The word frequency and consequent wordcloud of the total positive words and negative words for India (figure 1 and figure 2) respectively was calculated and presented through wordcloud. Similarly, frequency of positive words and negative words for USA were calculated and presented through wordcloud (figure 3, figure 4 and figure 5).

Fig. 1. Wordcloud India- Positive Words

Fig. 2. Wordcloud India- NegativeWords

Fig. 3. Wordcloud USA- Positive Words

Fig. 4. Wordcloud USA- Negative Words

Fig. 5. Wordcloud India-

Fig. 5. Wordcloud U S

Pandemic has created emotional distress Emotional distress during pandemic times plays havoc on the mind of people [9]. To further understand the sentiments, sentiment analysis for identifying the polarity of sentiments (positive, negative, or neutral) in the range of $-5 \leq v \leq 5$ (AFINN 2011) was calculated. Besides analysing the sentiment of reviews, we measured emotions of words in tweets The different emotion annotations for a target term were by determined by majority class of emotion intensities. The NRC emotion lexicon [21] was used to calculate the presence of ten basic emotions i.e.fear, anticipation, disgust, trust, anger, surprise, sadness, joy, surprise, anger, positive and negative. and their respective percentage. The result expressed broad sentiments presented in Table 1. In order to examine the difference in the emotional state of tweeters of two chosen countries independent sample t test was conducted. It is observed that out of ten emotions there is a significant difference on the four emotions between two groups i.e. India and USA (Table 2). These four emotions are fear, anger, sadness and negative. The tweet data reveals the negative emotions riding the minds of the Indian twitters as they feel fear (m=0.25, SD=0.19), anger (m=0.24, SD=0.20), sadness (m=0.18; SD=0.14) and negative (m=0.22; SD=0.16). Twitter, with 500 million users and 100 million tweets per day has become a valuable source of data [22]. By extracting and analyzing the sentiments of the tweets organizations, policymakers' marketers can seek insight into the pervading emotions and sentiments of the stakeholder in a particular context such as product, brand [23], new governance policy or the impact of pandemic virus [24]. The data collected for two major countries that have reported high number of corona cases gives some startling insights. It is found that among tweeters whose source or location was India have significantly high negative emotions (Table 2) and are more fearful, angry and sad. Although there is general pervasiveness of negative emotions around the globe due to unprecedented nature of pandemic but a closer look gives deep insight. Past researches provide evidence of perception of pandemic related negative emotions among Americans, but it is found to be more acute among those who have experienced COVID symptoms and have been diagnosed [25]. This can be attributed to American public health initiatives that are promoting safety through social distancing and minimizing transmission [24]. Better health expertise assists in ensuring effective pandemic response that curbs the curve of emotional distress [25].

Table I. NRC Sentiment Categorization

Table 1: NRC Sentiment Categorization				
	USA		India	
Sentiment	frequency	Percent	frequency	Percent
fear	1141	13.08	946	13.34
anticipation	676	7.75	671	9.46
disgust	516	5.91	226	3.19
trust	834	9.56	863	12.17
anger	662	7.59	423	5.96
sadness	1006	11.53	759	10.70
joy	403	4.62	364	5.13
surprise	489	5.61	252	3.55
positive	1213	13.90	1314	18.53
negative	1784	20.45	1275	17.98

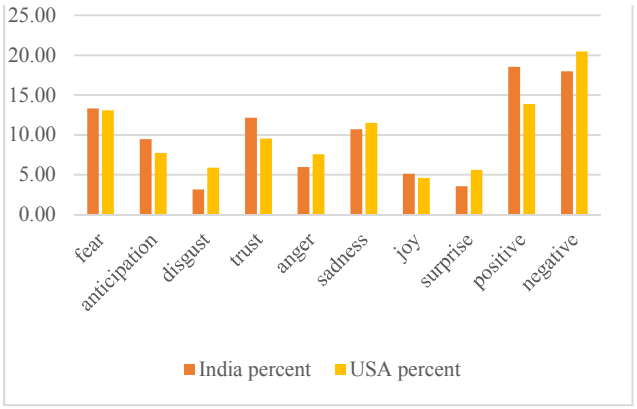

Figure 12 : USA and India-NRC Sentiments categorization

Table 2 : NRC Sentiment t table

	USA		India			
Sentiment	Mean	SD	Mean	SD	t value	Sig
Fear	0.21	0.18	0.25	0.19	4.85	0.00
anticipation	0.61	0.26	0.63	0.24	1.81	0.07
disgust	0.15	0.17	0.17	0.19	1.08	0.28
Trust	0.69	0.16	0.70	0.16	0.65	0.51
anger	0.20	0.21	0.24	0.20	3.04	0.00
sadness	0.17	0.15	0.18	0.14	2.11	0.03
Joy	0.82	0.13	0.83	0.12	1.08	0.27
surprise	0.52	0.28	0.52	0.31	-0.08	0.93
positive	0.74	0.14	0.73	0.14	-0.48	0.63
negative	0.20	0.15	0.22	0.16	4.70	0.00

V. Implications Limitations and Future Scope

The study gives a insight on the sentiments of the tweet user in two major countries. The pandemic had shaken the world and rising cases of infection and death had led people to emotional distress. However, this study which captures the tweet data in the month of September, that is seven months after WHO announced novel corona virus as pandemic. Emotions have evolved as the tweet natives in these months are empowered with better information on the precautionary measures. The study presents the current state of emotions using big data sample. The sentiment analysis assists the government and organization to gauge the present public

emotions and can guide them in taking strategic decisions related to business and governance.

However, this study is not devoid of limitations. The study only captures data for two main countries i.e. USA and India. A more detailed exploratory analysis of the individual tweets will give a better insight. Also, it is suggested that topical modelling along with sentiment analysis could be conducted for a more comprehensive study,

ACKNOWLEDGMENT

The paper is not funded and has no conflict of interest

REFERENCES

[1] S Li, Y Wang, J Xue, N Zhao, T Zhu, "The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users" International journal of environmental research and public health, vol17, no. 6, pp. 2032, 2020.

[2] SLD Restubog, ACG Ocampo, and L. Wang, "Taking control amidst the chaos: Emotion regulation during the COVID-19 pandemic". 103440, 2020.

[3] L. Cerbara , G. Ciancimino, M. Crescimbene, F. La longa, M.R. Parsi, A. Tintori, R. Palomba , "A nation-wide survey on emotional and psychological impacts of COVID-19 social distancing", Eur Rev Med Pharmacol Sci, vol.24, no.12, pp.7155-7163, 2020.

[4] Q. Chen, C. Min, W. Zhang, G. Wang, X. Ma, and R. Evans, "Unpacking the black box: How to promote citizen engagement through government social media during the COVID-19 crisis". Computers in Human Behavior, 106380, 2020.

[5] J. Van Bavel, K. Baicker, P.S. Boggio, V. Capraro, A. Cichocka, M. Cikara, ...and J. Drury, "Using social and behavioural science to support COVID-19 pandemic response". Nature Human Behaviour, pp.1-12, 2020.

[6] D.P. Aldrich, "Building resilience: Social capital in post-disaster recovery". 2012, University of Chicago Press

[7] H.W. Fischer III, "Behavior Response to Disaster. Response to Disaster Fact Versus Fiction and Its Perpetuation the Sociology of Disaster", 3rd edn. Lanham, MD: University Press of America, pp. 49-74, 2008

[8] G.A. Bonanno, "Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events?", American Psychologist, vol.59, no.1, pp 20-28, 2004.

[9] C.C. Sonn, and A.T. Fisher, "Sense of community: Community resilient responses to oppression and change". Journal of community psychology, vol.26, no.5, pp. 457-472, 1998.

[10] D. Mileti, "Disasters by design: A reassessment of natural hazards in the U.S". Joseph Henry Press, June 1999.

[11] S.K. Campbell and V.l. Butler, "Archaeological evidence for resilience of Pacific Northwest salmon populations and the socioecological system over the last~ 7,500 years" Ecology and Society, vol.15, no.1, 2010.

[12] S.L. Cutter, L. Barnes, M.Berry, C. Burton, E. Evans, E Tate, and J. Webb, "A place-based model for understanding community resilience to natural disasters", Global environmental change, vol.18, no.4, pp.598-606, 2008

[13] S.K.Harris, S.A. Smockand M. Tabor Wilkes, "Relapse resilience: A process model of addiction and recovery". Journal of Family Psychotherapy, vol. 22, no.3, pp. 265-274, 2011.

[14] M.E. Kruk, M. Myers, S.T. Varpilah, and B.T. Dahn, "What is a resilient health system? Lessons from Ebola". The Lancet, vol.385, no.9980, pp. 1910-1912, 2015.

[15] G.A. Bonanno, S.M. Ho, J.C. Chan, R.S. Kwong, C.K. Cheung, C.p. Wong and V.C. Wong, "Psychological resilience and dysfunction among hospitalized survivors of the SARS epidemic in Hong Kong: a latent class approach". Health Psychology, vol.27, no.5, pp659, 2008.

[16] G.D. Smith, F. Ng and W.H.C. Li, "COVID□19: Emerging compassion, courage and resilience in the face of misinformation and adversity". Journal of Clinical Nursing, vol.29, no.9-10, 1425,2020

[17] J. Qiu, B. Shen, M. Zhao, Z. Wang, B. Xie, and Y.Xu, "A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations". General psychiatry, vol.33, no.2, 2020.

[18] J.Wang, Z.Wang, X. Liu, X, Yang, M. Zheng, and X. Bai, "The impacts of a COVID-19 epidemic focus and general belief in a just world on individual emotions". Personality and Individual Differences, 110349, 2020

[19] A.D. Ong, C.S. Bergeman, T.L. Bisconti, and K.A.Wallace, "Psychological resilience, positive emotions, and successful adaptation to stress in later life. Journal of personality and social psychology, vol.91, no.4, pp. 730, 2006.

[20] L.Shanahan, A. Steinhoff, L. Bechtiger, AL. Murray, A. Nivette, U. Hepp, ... and M. Eisner, "Emotional distress in young adults during the COVID-19 pandemic: evidence of risk and resilience from a longitudinal cohort study". Psychological medicine, pp.1-10, 2020.

[21] S. Mohammad, and P.Turney "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon". in Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text pp. 26-34, June 2020.

[22] N. Pourebrahim, S. Sultana, J. Edwards, A. Gochanour, and S. Mohanty, "Understanding communication dynamics on Twitter during natural disasters: A case study of Hurricane Sandy". International journal of disaster risk reduction, vol.37, pp.101176, 2019.

[23] W. Read, N. Robertson, L. McQuilken, and A S Ferdous, "Consumer engagement on Twitter: perceptions of the brand matter". European Journal of Marketing, 2019.

[24] E. Chen, K. Lerman, E. Ferrara, "Covid-19: The first public coronavirus twitter dataset". arXiv preprint arXiv:2003.07372.2020.

[25] M W Gallagher, M J Zvolensky, L. J Long, A.H Rogers, and L. Garey, "The Impact of Covid-19 Experiences and Associated Stress on Anxiety, Depression, and Functional Impairment in American Adults". Cognitive Therapy and Research, pp. 1-9, 2020.