Uncovering COVID-19 conversations: Twitter insights and trends

Selim Molla 1, *, Ehsan Bazgir 2, S M Mustaquim 3, Iqtiar Md Siddique 4 and Anamika Ahmed Siddique 4

1 Department of Computational Science, University of Texas at El Paso, US.
2 Department of Electrical Engineering, San Francisco Bay University, Fremont, CA 94539, US.
3 Department of Mathematical Science, University of Texas at El Paso, US
4 Department of Mechanical & Aerospace Engineering, University of Texas at El Paso, US.

Publication history: Received on 02 December 2023; revised on 07 January 2024; accepted on 10 January 2024

Abstract

In this paper, we delve into the public discourse surrounding COVID-19 on Twitter to unearth the collective sentiments, concerns, and spread of information during the pandemic. By leveraging a dataset of relevant tweets and corresponding ISO country codes, our analysis will map out the geographical and digital landscape of these conversations. The significance of this work lies in its potential to inform public health strategies, shape policymaking, and contribute to social research on crisis communication. Stakeholders ranging from health officials to the public have a vested interest in understanding the contours of this dialogue. Our objective is to craft a data-driven narrative through visualizations that reveal how the world engages with the pandemic on the digital front, providing actionable insights into global and local responses to COVID-19 using Machine Learning techniques.

Keywords: COVID-19; Machine Learning; Twitter; Word Cloud.

1. Introduction

Social media platforms, such as Twitter and Facebook, have emerged as influential mediums for individuals to articulate their emotions and responses in relation to past or ongoing events. The COVID-19 pandemic has not only reshaped our daily lives but also revolutionized the way information and opinions are disseminated and shared globally. Amidst the unprecedented challenges posed by the virus, social media platforms like Twitter have become vital channels for individuals to express their thoughts, concerns, and emotions. These platforms have evolved into dynamic forums where discussions, debates, and information exchange related to COVID-19 are constantly unfolding. The insights gleaned from this research are poised to offer substantial benefits to a diverse range of stakeholders. Public health bodies, policymakers, and academics all stand to gain invaluable knowledge from understanding global conversations and regional sentiment variances. These insights can inform and guide decision-making processes, enabling the crafting of health messaging and policies that resonate with the public’s voice and effectively address the pressing issues highlighted through social media discourse during the ongoing pandemic. Twitter stands out as a prominent platform where users employ concise tweets, previously limited to a maximum of 140 characters and currently extended to 280 characters, to encapsulate their sentiments and reactions in a succinct manner. This characteristic renders Twitter an exceptional medium for comprehending the feelings and reactions of individuals towards a specific event, surpassing the analytical capabilities of other social media platforms. Considering this, researchers have dedicated their efforts towards developing visualization tools that facilitate the exploration and comprehension of individuals’ tweets from various perspectives. For instance, numerous visualization tools have been designed with a primary focus on showcasing sentiment analysis of tweets, as exemplified by [1-10]. Conversely, alternative visualization tools concentrate on the analysis and exploration of tweet data based on its geo-spatial information, as demonstrated by [11-17]. The COVID-19 outbreak initially surfaced in news reports during the month of December 2019. Nonetheless, the
gravity of the pandemic only began to manifest itself in the early months of 2020, when it garnered heightened attention and individuals began to comprehend the severity of the situation. As worldwide lockdowns were implemented in response to the pandemic’s rapid spread, individuals increasingly turned to social media platforms to express their sentiments and reactions towards COVID-19.

One study proposed a framework that dynamically identifies key topics with labels from COVID-19 tweets using Latent Dirichlet Allocation (LDA) generated topics [18]. Another study examined linguistic and visual communication in UK public health agency tweets related to COVID-19. It found that the combination of linguistic and visual representations in these tweets affected the overall effectiveness of public health communications [19]. Additionally, a study analyzed the sentiments of Indian people towards COVID-19 vaccines on Twitter. It found that most tweets reflected positive sentiments and identified the most frequently occurring terms used in these tweets [20]. Furthermore, a study used visualization techniques to explore people’s tweets about the COVID-19 pandemic over time, providing insights into their feelings and reactions [21]. Lastly, a study proposed an IPSH model for predicting COVID-19 tweet content using parts of speech tags and high-frequency words, which was found to be more efficient than baseline machine learning models [22]. Ahmmed et al. (2023) describes how brain tumor imaging has been detected through the deep learning algorithm and we have plan to extend our current research to this stage in future [24]. Ullah, et al. (2023) try to find out the best scenarios for a production floor that considers covid 19 protocol as well for their industry [31]. They also apply this covid system problem into manufacturing operation that used simulation as well [32]. Moreover, Hossain at. el. (2023) details in his research regarding electricity generation from a moving vehicle which can be the further research proposal from our finding [27]. Supplier selection was the focus, but the author uses comparative weightage method using C algorithm which is important for this research [28,29]. In this study, Fayshal et al.(2023) discusses how human health risk have significant effect on humans and our research will take this into consideration for covid 19 [30]. So, from the above all related works it is undoubtedly true that in this post covid period, to consider the COVID protocol in the daily work is an important task and this research will show how much spread the covid across the various zone in the world from which one can impose restriction for the measurements. In this endeavor, we have harnessed the power of data analytics to not only map sentiments but also identify influential communication channels and highlight prevalent concerns. Nevertheless, it is important to acknowledge the scale of our undertaking, as the sheer volume of Twitter data is vast and ever evolving. Therefore, while we strive to provide comprehensive insights, we must also recognize the inherent limitations posed by the unceasing flow of information in the digital sphere. Molla et al. (2023) describes how the medical textiles sector is influenced by covid-19 protocol and we have future research on this category [34].

This study serves as a vital exploration into the world of COVID-19 conversations on Twitter, offering a nuanced understanding of the ongoing discourse and trends. It is our hope that the findings and analyses presented in the following pages will contribute significantly to our collective knowledge of public sentiment during these challenging times and, in turn, aid in the development of more informed, effective, and responsive strategies for addressing the pandemic’s multifaceted impact.

2. Methods / Data Understanding

2.1. Data Collected

The data collected for this study comprises tweets related to COVID-19 sourced from Twitter, along with a dataset containing Wikipedia ISO country codes. The dataset encompasses individual tweets, user information, and corresponding country codes. This comprehensive collection offers a broad perspective on the global discourse surrounding the COVID-19 pandemic [23].

2.2. Data Description

Data provides a description of individual tweets, user information, and country codes, which together offer a broad view of the global discourse on the pandemic.

2.3. Data Manipulation

Cleaning (handling missing values, correcting data types), merging (associating tweets with standardized country codes), and transforming (extracting hashtags, processing dates and times for temporal analysis).

- Cleaning (handling missing values, correcting data types), merging (associating tweets with standardized country codes), and transforming (extracting hashtags, processing dates and times for temporal analysis).
Data to conduct the study: This data answers the research questions about public sentiment and information dissemination during the COVID-19 pandemic.

Figure 1 Distribution of the top 10 sources in pie chart

2.4. Data to conduct the study.

The purpose of utilizing this data is to address research questions concerning public sentiment and information dissemination during the COVID-19 pandemic. By diving into the subjects using specific hashtags, examining geographical distribution, evaluating sentiments expressed in tweets, and identifying trends over time, the study aims to gain valuable insights into the public discourse surrounding the pandemic on a global scale. Through these analyses, the study seeks to contribute to a comprehensive understanding of how information flows and sentiments evolve in the context of a global health crisis.

2.5. Visualizations

2.5.1. Bar Graph

Representation of the distribution of tweet sources, which helps in identifying where the most active discussions originate from.

Figure 2 Top User from India: 1.96% and the second Top is USA

2.6. Word-cloud

This visualization highlights the most prevalent words within the tweets, especially from high-volume regions like India, which aids in discerning the dominant themes in the discourse. From the below figure most frequently quoted words are coronavirus and in India it is shown that COVID19 is the most used from the word cloud. Rahman et al (2023)
interprets the bitcoin miners who consider the Covid protocol for their sitting arrangement during mining process [25,33]. Another study reveals that when they try to consider process for their jute mills covid protocol was the significant factor [26].

![Prevalent words in tweets](image1)
![Prevalent words in tweets from India](image2)

**Figure 3 Prevalent words from tweets and from India**

### 2.7. Choropleth map's

The geographic distribution of words is visualized in a choropleth map, where the density or frequency of certain terms used in tweets is represented by varying colors across geographical locations. This allows for a visual interpretation of the spatial aspect of the data, highlighting regions that are more active or have different focuses in the discourse on the pandemic. The choropleth map uploaded could be described as "Fig 4" in this section. It likely represents the geographic distribution of tweet volume, or the frequency of specific words related to COVID-19, as indicated by the color gradient. The map would be interpreted according to the legend, which appears to correlate color intensity with the number of tweets. This form of visualization is a powerful tool for presenting geographic variations in data and can be particularly illustrative of how different regions contribute to the global conversation on Twitter regarding COVID-19.

![Figure 4 Geographic location for the words in choropleth](image3)

### 2.8. Evaluation

For stakeholders such as public health officials and policymakers, the results might indicate areas where public concern is highest, misinformation is spreading, or where more communication is needed. For researchers, these insights could signal trends worthy of deeper study, such as the relationship between public sentiment and compliance with health guidelines.

- **Public Health Communications**: Tailor messaging to address prevalent concerns and misinformation identified in tweets.
- **Policy Formulation**: Consider the sentiment and topics of discussion when creating or adjusting COVID-19 related policies.
- **Engagement Strategies**: Engage with key influencers and leverage popular hashtags to improve public health message dissemination.
- **Further Research**: Conduct longitudinal studies to see how sentiments and topics change over time, which could inform future crisis communication strategies.
The visualizations would have served as a foundation for these insights and recommendations, providing stakeholders with actionable data to inform their decisions and strategies.

3. Conclusions

This study analyzes Twitter data to grasp public sentiment on COVID-19, identifying key topics and dissemination patterns. In conclusion, this study has undertaken a comprehensive analysis of Twitter data to gain valuable insights into public sentiment regarding COVID-19. By examining a vast corpus of tweets and utilizing ISO country codes, we have successfully mapped sentiments and information dissemination patterns. This effort has allowed us to uncover not only the prevailing concerns but also the influential communication channels shaping global conversations and regional sentiment variances. The significance of these findings cannot be overstated, as they offer crucial guidance to public health bodies, policymakers, and academics. These insights enable more informed decision-making, helping to craft effective health messaging and policies that resonate with the public's voice and address the pressing issues highlighted through social media discourse during the pandemic. However, it is important to acknowledge a limitation of our research, which stems from the sheer volume of data. Despite our best efforts, we were unable to analyze every tweet in the dataset. Nonetheless, the depth and breadth of our analysis provide a robust foundation for understanding the dynamic landscape of COVID-19 conversations on Twitter and offer a valuable resource for future research and action in the field of public health communication and policy development.

Future Work

To expand the COVID-19 Twitter analysis project, incorporating diverse data like news articles and health reports, alongside advanced analytics like sentiment and network analysis, could yield deeper insights. This would enable tracking sentiment trends, topic shifts, and misinformation spread, and could measure influencer impact and the effectiveness of health responses. Access to broader data and advanced tools for processing natural language would be necessary for this enhanced analysis.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References


