TweetCOVID: A System for Analyzing Public Sentiments and Discussions about COVID-19 via Twitter Activities

JOLIN SHAYNN-LY KWAN, Singapore University of Technology and Design, Singapore

KWAN HUI LIM, Singapore University of Technology and Design, Singapore

The COVID-19 pandemic has created widespread health and economical impacts, affecting millions around the world. To better understand these impacts, we present the TweetCOVID system that offers the capability to understand the public reactions to the COVID-19 pandemic in terms of their sentiments, emotions, topics of interest and controversial discussions, over a range of time periods and locations, using public tweets. We also present three example use cases that illustrates the usefulness of our proposed TweetCOVID system.

CCS Concepts: • **Human-centered computing** → **Social networking sites**; • **Information systems** → **Web and social media search**; *Social networks*.

Additional Key Words and Phrases: COVID-19, Twitter, Sentiment Analysis, Topic Modelling

1 INTRODUCTION

The COVID-19 pandemic has infected millions and claimed the lives of hundreds of thousands around the world. In addition, COVID-19 has caused widespread economic impact, resulting in extreme levels of unemployment across all industries. In turn, these medical and economical impacts have created immense emotional stress on the global population. With the ongoing pandemic, it is important to understand the ever-changing impact of COVID-19 as new regulations and developments take place over the world.

Towards this effort, we present TweetCOVID, a system for understanding the impact of COVID-19 on the general public in terms of sentiments, emotions, topics and controversial discussions. TweetCOVID utilizes publicly available tweets, alongside sentiment analysis (levels of positivity and negativity), emotion detection (anger, fear, sadness, disgust, surprise, anticipation, trust, and joy), topic modelling using Latent dirichlet allocation (LDA), controversial term tracking and various temporal and spatial analysis.

1.1 Related Work

Twitter has been used for various applications, such as online misbehaviour detection [3, 5], event tracking [1, 4, 9], activity recommendation [10, 17], etc. While there has been work using Twitter for COVID-19 related studies, many of these works focus on a single region [15, 16] or for a specific single purpose, e.g., misinformation [6, 14]. In contrast, our

proposed TweetCOVID system provides a comprehensive set of functionalities to study sentiments, emotions, topics, controversial terms across various timeframes and locations. This holistic set of capabilities allow the user to study a flexible range of questions, e.g., what are the public sentiments in a certain country towards the implementation of a lockdown and what are the following public discussions?

2 SYSTEM ARCHITECTURE AND FUNCTIONALITIES

Our proposed system comprises the main components for data collection/processing, general analytics, sentiment/emotion analysis, topic modelling, controversy tracking and visualisation. These components are further elaborated next.
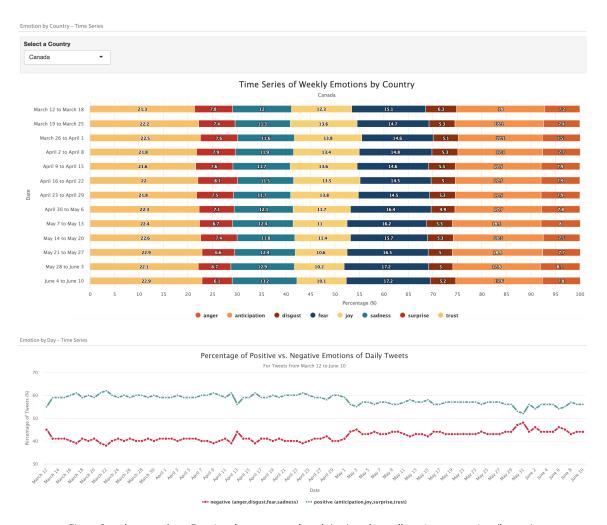

Fig. 1. Sample screenshots: Emotions by country and week (top), and overall sentiment over time (bottom).

- **Data Collection/Processing**: This component allows the user to either load in existing Twitter datasets or retrieve datasets via the Twitter API. After loading/collecting these dataset, the standard processing steps of

language filtering, tokenization, stemming and lowercase conversion are applied to the dataset before the subsequent analysis.

- **General Analytics**: This component shows the general statistics and distribution of the dataset. For example, a user can view the number of tweets posted in specific time periods and/or at geographic locations to understand how frequent the discussion about COVID-19 is in specific areas.

- **Sentiment/Emotion Analysis**: This component shows the general sentiments and specific emotions associated with COVID-19 related discussions. We use the AFINN lexicon [12] for sentiment analysis that allows us to determine the positivity and negativity of specific tweets, which we further aggregate for the overall COVID-19 discussions. For specific emotions, we use the NRC Word-Emotion Association Lexicon [11] that allows us to determine the extent which a tweet displays the eight primary emotions based on Plutchik theory [13]. Figure 1 shows two screenshots of the sentiment and emotion analysis components.

- **Topic Modelling**: This component displays the various topics that were discussed over time during the course of the pandemic. We used the popular LDA [2] on the set of weekly tweets to determine the main topics discussed on a weekly basis. LDA represents tweets as a random mixture over latent topics and each topic is in turn represented by a set of words. Figure 2 (top) shows a screenshot of this topic modelling component.

- **Controversy Tracking**: This component tracks and identifies tweets that include various controversial terms, such as "wuhan virus", "chinese virus", "kung flu", etc. These tweets are controversial in their content and designed to incite hatred towards a specific group of people. Figure 2 (bottom) shows a screenshot of this controversy tracking component.

- **Visualisation**: This component provides a wide range of visualisation techniques to best display the processed results from the above-mentioned components. In addition, a user is able to drill-down and present these analysis across different time periods and locations.

Figures 1 and 2 shows various example screenshots of our TweetCOVID system. For a more interactive demonstration of these and additional features, we refer interested readers to https://youtu.be/ut2EzLeuFwo for a video demonstration.

3 EXAMPLE USE CASES / DEMONSTRATION

Using the set of functionalities provided by TweetCOVID, we can utilize the system for various use cases, which we describe with three examples. For a more detailed discussion of the main findings, we refer interested readers to our earlier paper [8] that elaborates on various research questions and findings related to the COVID-19 pandemic.

Use Case 1: Understanding public sentiments and emotions in relation to global events. Using our system, we observe that public sentiments were noticeably more negative on 12 March 2020 as compared to the weeks before and after. Further analysis shows that anxiety and fear levels were heightened after the WHO Director-General declared COVID-19 as a pandemic on 11 March 2020. Thereafter, there was a corresponding increase in discussions relating to the pandemic and the potential impact on the users' daily life.

Use Case 2: Detecting adjacent topics during the course of the pandemic. Apart from the COVID-19 pandemic, there were various global incidents that happened concurrently with the pandemic. Our TweetCOVID system was able to pick up such incidents in the form of discussion topics. For example, we detected the topic of the Baltimore Shooting over 12 to 18 March 2020 and how certain users were linking this incident back to COVID-19.

Use Case 3: Understanding the use of controversial terms during the pandemic. Our proposed system also tracks the use of controversial terms, such as "wuhan virus", "chinese virus", "kung flu", and the context they were used

Top 5 Topics Deduced From Keywords Identified Through LDA

Select a Week

March 12 to March 18 ▾

Call to Action	Baltimore Shooting	Social Networking in Quarantine	Cancellation of Events / Closure of Schools & Workplaces	United Kingdom's Approach to COVID-19
will	troll	profile	vlog	ngo
have	baltimore	eachother	invisible	wellness
get	century	selfie	respectful	mail
people	paramedic	welfare	australiangp	meter
can	morocco	almighty	ksleg	bojo
time	politicize	facilitate	teleworking	cent
good	universe	shed	receive	discharge
now	vet	golden	sanitary	hesitate
just	murder	neighbour	random	sufficient
need	constituent	presence	thermometer	tesco

Note: highlighted words are the words used to derive the names of the topics

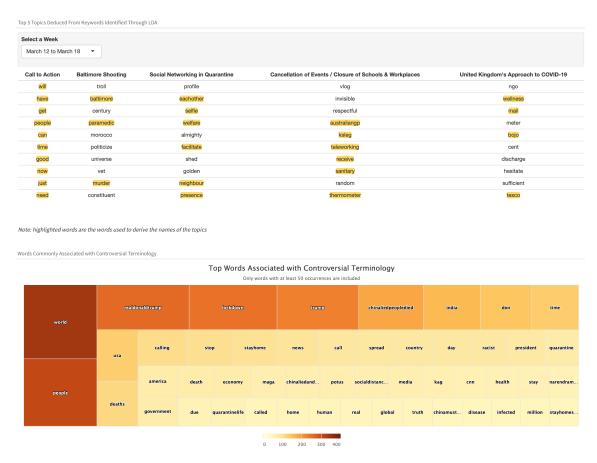Words Commonly Associated with Controversial Terminology

Fig. 2. Sample screenshots: Discussion topics by week (top), and top words associated with controversial terms (bottom).

in. We found that more than half of these usage were from the USA, with a majority from political figures during March 2020. Additionally, the racist nature of the controversial terms can be seen through the co-occurrence of terms such as "chinaliedpeopledied", "racist" and "chinamustexplain". These terms suggest misplaced blame upon people of Asian descent for the pandemic and further perpetuates racism in countries.

4 CONCLUSION

In this paper, we present a demonstration of the TweetCOVID system for understanding the impact of COVID-19 on the general public using publicly available tweets. TweetCOVID offers a range of functionalities, including data collection/processing, general analytics, sentiment/emotion analysis, topic modelling, controversy tracking and visualisation. We also demonstrate three use cases that our system can be used for, including understanding sentiments/emotions, detecting topics alongside COVID-19 and understanding controversial terms used. Future work will include exploring other topic models that are relevant for short text [7].

ACKNOWLEDGMENTS

REFERENCES

[1] Hamed Abdelhaq, Christian Sengstock, and Michael Gertz. 2013. Eventweet: Online localized event detection from twitter. *Proc. of the VLDB Endowment* 6, 12 (2013), 1326–1329.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *JMLR* 3, Jan (2003), 993–1022.

[3] Despoina Chatzakou, Nicolas Kourtellis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Athena Vakali. 2017. Mean birds: Detecting aggression and bullying on twitter. In *Proc. of WebSci.* 13–22.

[4] Carmela Comito, Agostino Forestiero, and Clara Pizzuti. 2019. Bursty event detection in Twitter streams. *TKDD* 13, 4 (2019), 1–28.

[5] Paolo Fornacciari, Monica Mordonini, Agostino Poggi, Laura Sani, and Michele Tomaiuolo. 2018. A holistic system for troll detection on Twitter. *Computers in Human Behavior* 89 (2018), 258–268.

[6] Anatoliy Gruzd and Philip Mai. 2020. Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data & Society* 7, 2 (2020).

[7] Elias Jónsson and Jake Stolee. 2015. An evaluation of topic modelling techniques for twitter.

[8] Jolin Shaynn-Ly Kwan and Kwan Hui Lim. 2020. Understanding Public Sentiments, Opinions and Topics about COVID-19 using Twitter. In *Proc. of ASONAM.*

[9] Kwan Hui Lim, Sachini Jayasekara, Shanika Karunasekera, Aaron Harwood, Lucia Falzon, John Dunn, and Glenn Burgess. 2018. RAPID: Real-time Analytics Platform for Interactive Data Mining. In *Proc. of ECML-PKDD.* 649–653.

[10] Sreekanth Madisetty. 2019. Event Recommendation using Social Media. In *Proc. of ICDE.* 2106–2110.

[11] Saif Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In *Proc. of SemEval.*

[12] Finn Årup Nielsen. 2011. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In *Proc. of MSM.* 93–98.

[13] Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion.* 3–33.

[14] Monica Stephens. 2020. A geospatial infodemic: Mapping Twitter conspiracy theories of COVID-19. *Dialogues in Human Geography* 10, 2 (2020), 276–281.

[15] Yue Su, Jia Xue, Xiaoqian Liu, Peijing Wu, Junxiang Chen, Chen Chen, Tianli Liu, Weigang Gong, and Tingshao Zhu. 2020. Examining the impact of COVID-19 lockdown in Wuhan and Lombardy: A psycholinguistic analysis on Weibo and twitter. *Intl. J. of Environmental Research and Public Health* 17, 12 (2020), 4552.

[16] Stefania Vicari and Maria Francesca Murru. 2020. One Platform, a Thousand Worlds: On Twitter Irony in the Early Response to the COVID-19 Pandemic in Italy. *Social Media+ Society* 6, 3 (2020).

[17] Jia Wang, Yungang Feng, Elham Naghizade, Lida Rashidi, Kwan Hui Lim, and Kate Lee. 2018. Happiness is a choice: sentiment and activity-aware location recommendation. In *Proc. of The Web Conference.* 1401–1405.