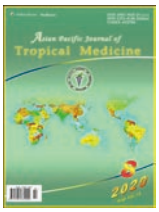


Letter to Editor

Asian Pacific Journal of Tropical Medicine

journal homepage: www.apjtm.org

doi: 10.4103/1995-7645.279651

Impact Factor: 1.94

Using twitter and web news mining to predict COVID-19 outbreak

Kia Jahanbin[1], Vahid Rahmanian[2]✉

[1]*Information Technology, Islamic Azad University Branch of Kerman, Iran*

[2]*Zoonoses Research Center, Jahrom University of Medical Sciences, Jahrom, Iran*

On January 9, 2020, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), formerly known as 2019-nCoV, was declared the causative agent in 15 of the 59 hospitalized patients in Wuhan, Hubei Province, causing great concern: this new coronavirus has 70% genetic association with SARS and is a subspecies of Sarbecovirus. The virus is temporarily named the 2019-nCoV virus[1] and the Coronavirus Study Group has nominated the virus as SARS-CoV-2[2].

In January 2020, more positive cases from other countries such as Thailand, Japan, South Korea, and the United States of America were reported by January 20, 2020, and the transmission of individual-to-health care, further complicated the situation[3].

Coronaviruses are zoonotic, meaning they are transmitted between animals and people, but the ways in which it is transmitted, animal reservoirs, prophylaxis, and precise clinical manifestations requires more investigation. There is currently no vaccine and appropriate treatment for COVID-19, so a high index of clinical suspicion and inquiring about the history of travel and contact from patients with fever and respiratory symptoms play a critical role in the prevention and control of the disease[4].

On a daily basis, a large number of Websites and online social media produce a large amount of data in a variety of fields such as technology, medicine, history, political and social news, arts and other fields. Analyzing and classifying these data leads to the production of knowledge and nowadays, it has attracted the attention of many researchers[5].

Web news mining is one of the most significant tools and the subset sciences "Big Data" in social networking. A web news mining-based automatic system can monitor, evaluate, and categorize news, which, in addition to managing news articles, it is also applied in the field of advisory systems[6].

Social networks fall into six groups as follows[7]: 1. Micro-blogging platforms: such as twitter; 2. Blogging platforms: such as WordPress and Blogger; 3. Instant messaging Apps: such as WhatsApp and Telegram; 4. Networking platforms: such as Facebook and LinkedIn; 5. Software elaboration platform: such as GitHub; 6. Photo/video sharing platforms: such as Instagram and YouTube.

The Twitter social networking is a micro-blogging platform considered by researchers as a result of useful applications. There are over 320 million active subscribers on the social network, which daily generates approximately 6 million tweets containing instant news and comments; due to the wealth of information and their easy access. Twitter has extensive applications, such as the predicting a political process, investigating the effectivity of a product, monitoring the events pertaining to the health and hygiene[8]. Approximately 23% of Twitter subscribers are adults and on a daily basis, a total of approximately 500 million Tweets are broadcasted each day[9].

In the model presented in this study, unstructured data on a novel coronavirus (2019-nCoV) are extracted from Twitter and then subjected to text cleaning, so-called screening or filtering, and finally classification operations. Since the focus is on real-time programming, this model is implemented using a fuzzy rule-based evolutionary algorithm called Eclass1-MIMO.

One of the most effective ways to prevent and control epidemics is to monitor and track the news and social networks about the spread of infectious diseases. In this study, the FAMEC method was used

✉To whom correspondence may be addressed. E-mail: rahmanian.vahid@ut.ac.ir

How to cite this article: Jahanbin K, Rahmanian V. Using twitter and web news mining to predict COVID-19 outbreak. Asian Pac J Trop Med 2020; 13(8): 378-380.

Article history: Received 19 February 2020 Revision 25 February 2020
Accepted 26 February 2020 Available online 2 March 2020

to send an alert message to surveillance systems for timely detection outbreaks of the COVID-19.

The FAMEC method has four main phases as follows:

1. Clearing and integrating data and extracting vocabulary; 2. Web and tweet crawling; 3. Applying fuzzy rules and storing data using fuzzy classifier. 4. Visualizing and sending messages.

The visualization component of the suggested method aims to assist in real-time monitoring and tracking of the beginning and spread of outbreaks, which can greatly contribute to the effectiveness of public health surveillance systems in this area.

Initially, during the period between Dec. 31 2019 and Feb. 6 2020, 2019-nCoV (COVID-19) tweets were extracted from the Twitter social network and stored in the relevant database. The collected database contained 364 080 tweets from 179 534 users. 21 805 371 users who have re-tweet or like these posts and 52 837 975 554 times these posts have been viewed by users. The main hashtags about novel coronavirus were #corona, #ncov, #wuhan, #china, #2019-nCoV, #virus, #corona virus china, #coronavirus outbreak, wuhan virus.

Figure 1 shows the results obtained from the monitoring of a novel coronavirus (2019-nCoV) related news in the study period, which are associated to 364 080 tweets from 179 534 users. The most Tweets about the coronavirus have been from the US (42.1%), China (13.0%), Italy (11.8%) and Australia (6.6%). This is consistent with the report of the cases which was obtained from the WHO[10]. In this study, a new method based on fuzzy algorithm was applied for evolving of the TSK of mining, monitoring, storage and visualization of news and tweets about preparing our COVID-19. To execute the method, more than 364 080 clean and integrated tweets and news were then categorized using the Eclass1-MIMO method and finally viewed in real time on the world map.

In the recent years, a significant number of researchers have been working on categorizing, clustering, analyzing emotions, thinking and developing recommenders based on social data, but most of these works have focused on either news websites or Twitter.

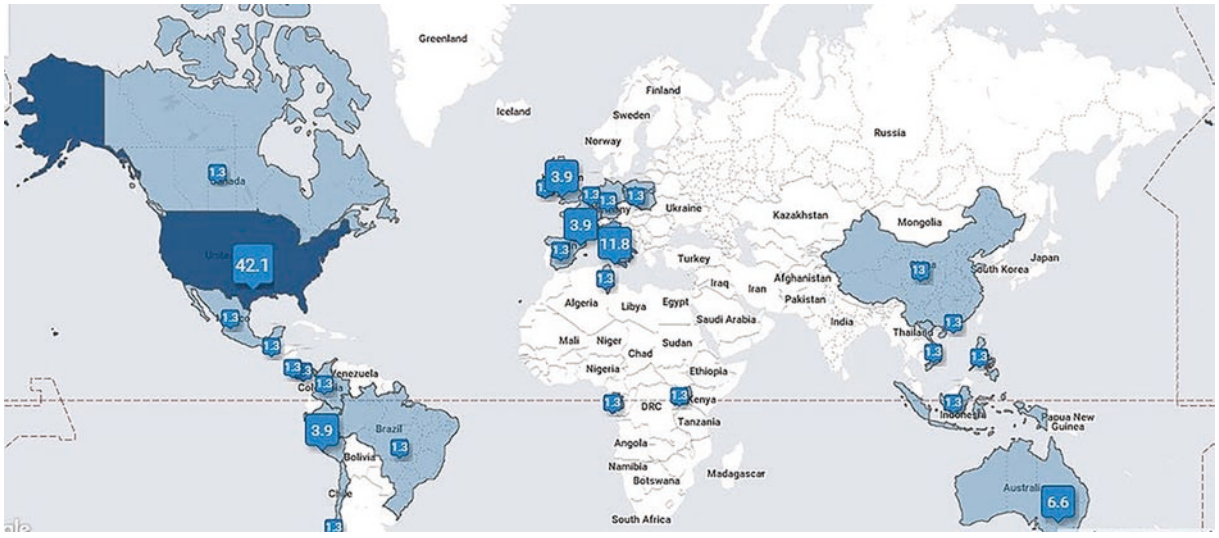
The evolving fuzzy algorithm with the Eclass1-MIMO method was used in the study of Iglesias for classifying six areas of knowledge, health, technology, sports, arts and commerce[5]. Also, Jahanbin *et al.* used web news mining in infectious disease surveillance systems to timely diagnose epidemics[8].

The geographical origins of tweets posted about COVID-19 were found to be consistent with the formal WHO report about incidence cases of COVID-19 during the study period. This reflects the efficacy of the suggested method to monitor and track this infection. The limitation of the proposed method is that it cannot be used to monitor and track infectious diseases in regions with poor or no access to social networks such as Twitter and Facebook. Also, as the language of processing the tweets in this study was English, the results may be affected by the processing language.

In conclusion, due to the revolutionary development of the social networks, using the web news mining of these network used by each community, the geographical and demographical of the users can be identified accurately. This is due to the fact that these network report easily statistical data with the most comments, photos, videos, *etc*. on COVID-19. This helps to predict morbidity rates in each region, and bring attention of policy-maker in the health care systems to purposefully implement educational programs in the regions where exposed to higher risks. Finally, this can help to reduce the incidence case and even mortality in communities.

Conflict of interest statement

The authors declare that there is no conflict of interest.



Figuer 1. Monitoring of geographical distribution of the tweets about COVID-19 between 31/12/2019 and 6/02/2020.

Acknowledgment

The authors would like to thank to the instructors of the online course "Machine Learning for Data Science and Analytics" provided by Columbia University for giving us better insight into the area of data and text mining.

Authors' contributions

VR, and KJ conceived and designed the study. VR, and KJ were responsible for literature search and screening. KJ were responsible for data collection and analyses. VR, KJ, contributed to data interpretation. KJ drafted the manuscript and VR, critically revised the manuscript.

References

[1] Nishiura H, Jung SM, Linton NM, Kinoshita R, Yang Y, Hayashi K, et al. The extent of transmission of novel coronavirus in Wuhan, China, 2020. *J Clin Med* Jan. 24 2020. doi: 10.3390/jcm9020330.

[2] Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. Severe acute respiratory syndrome-related coronavirus: The species and its viruses-a statement of the coronavirus study group. *bioRxiv* 2020.02.07.937862; doi: https://doi.org/10.1101/2020.02.07.937862.

[3] Majumder M, Mandl KD. Early transmissibility assessment of a novel coronavirus in Wuhan, China. *SSRN* Jan. 26 2020. doi: http://dx.doi.org/10.2139/ssrn.3524675.

[4] World Health Organization. *Coronavirus 2020*. [Online]. Available from: https://www.who.int/health-topics/coronavirus [Accessed on 10 February 2020].

[5] Iglesias JA, Tiemblo A, Ledezma A, Sanchis A. Web news mining in an evolving framework. *Inf Fusion* 2016; **28**: 90-98.

[6] Guellil I, Boukhalfa K. Social big data mining: A survey focused on opinion mining and sentiments analysis. In: *Conference of ISPS 2015: 12th International Symposium on Programming and Systems*. Algiers: IEEE; 2015. doi:10.1109/ISPS.2015.7244976.

[7] Ravindran SK, Garg V. *Mastering social media mining with R*. Mumbai: Packt Publishing Ltd; 2015.

[8] Jahanbin K, Rahmanian F, Rahmanian V, Sotoodeh Jahromim A. Application of Twitter and web news mining in infectious disease surveillance systems and prospects for public health. *GMS Hyg Infect Control* 2019; **14**: 1-12.

[9] Duggan M, Ellison NB, Lampe C, Lenhart A, Madden MJPRC. Social media update 2014. *Pew Res Center* 2015; **19**(9): 1-17.

[10]World Health Organization. *Novel coronavirus (2019-nCoV) situation reports. 2020*. [Online]; Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/. [Accessed on 10 February 2020].