

Sentiment Analysis and Opinion Mining about COVID-19 vaccines of Twitter Data

KIA JAHANBIN¹, VAHID RAHMANIAN², NADER SHARIFI², FERESHTEH RAHMANIAN³

¹Department of Information Technology, Yazd university, Yazd, Iran

²Research Center for Social Determinants of Health, Jahrom University of Medical Sciences, Jahrom, IR Iran

³Department of Information Technology, Islamic Azad University Branch of Kerman, Iran

Correspondence to Dr. Vahid Rahmadian, MPH, Ph.D. in Epidemiology, Email: vahid.rahmani1392@gmail.com, Tel: +989175985204

Dear Editor,

Social networking sites like Twitter have been a prevalent option for individuals to express their opinions, report real-life events and provide visions into what is happening around the world. In the epidemic of COVID-19, individuals have used Twitter to spontaneously share data visualization from news media and government agencies and send visualized data images they have generated separately¹.

Microblogging websites have evolved to develop a source for a variety of information. This is because of the nature of the microblogs in which individuals post real-time messages about their views on various topics, discuss current issues, complaints, and express their sentiment about events in everyday life².

Sentiment analysis is the field of research that evaluates public's opinions, feelings, assessments, attitudes, and feelings from writing language. It is one of the active areas of natural language processing and is extensively studied in data mining, web mining and text mining³.

This research has expanded into management and social sciences outside of computer science because of its impact to the whole society. The rising importance of sentiment analysis is evolving as social network evolves, such as reviews, forum discussions, blogs, microblogs, Twitter and social media⁴.

Microblogging data such as Twitter, where users messages real-time feedback and statements almost about "everything", make newer and various challenges. Some of the studies on sentiment analysis of Twitter data were conducted by Go et al⁵, and Pak and Paroubek⁶. Go et al. used distance learning to collect the sentiment data. They used tweets ending in positive emotions like ":" ":-)" as positive and negative emotions like ":(":(:" as negative⁵.

Another study for sentiment analysis on Twitter data was reported by Barbosa and Feng⁴. They applied of syntax features of tweets like a retweet, hashtags, link, punctuation and exclamation marks for the sentiment analysis.

In this work, we investigated one of the famous microblogs in Twitter to find out how the others think about the vaccines of new Coronavirus. We have implemented a fuzzy rule-based evolutionary algorithm called Eclass1-MIMO⁷ and then built three classify "tweets" into positive (hopeful and optimism to control this epidemic), negative (frustration, stress, and anxiety), and neutral (apathetic and trivial for the user) sentiment.

The search was implemented using the keywords as follows: coronavirus, ncov, COVID-19, virus, 2019-nCoV, vaccine, immunization, Pfizer corona vaccine, Moderna vaccine, AstraZeneca vaccine, coronavirus outbreak, Epidemic, only or in combination, in English language. In total, 1,126,137 tweet about vaccine of covid-19 were extracted between 01 December and 30 of December 2020. Figure 1 shows the results obtained from the monitoring of a vaccine of covid-19 related news within the period of this study.

Classifying "tweets" into positive sentiment means that the people who discovered the Covid-19 vaccine has given them a sense of salvation, hope for an end to the Corona pandemic and a return to pre-pandemic life. This group tends to be vaccinated if the vaccine is available. Classifying "tweets" into negative sentiment means that people feel cowardly and distrustful of the Covid-19 vaccine and are reluctant to vaccinate if the vaccine is available. Classifying "tweets" into neutral sentiment group means that people did not feel happy or scared about the news of the discovery of the Covid-19 vaccine.

The result of this study showed that 591053 tweet (52%) were in positive sentiment, 382431 (34%) neutral sentiment and 152653 (14%) negative sentiment about vaccine of covid-19.

The question, is that why's people, despite have knowledge that COVID-19 is spread in all the world, and lead to hospitalization or deaths, but they do not welcome vaccination and are afraid to do so (negative and neutral groups).

The key principle is that each individuals and communities should identify and practice good lifestyles, maintain health and try to prevent the spread of virus by a proper behavior training, and consider the role of health education and social marketing^{8,9}.

Social marketing in public health, the use of marketing to design and implement programs to promote socially beneficial behavior change, has grown in popularity and usage within the public health community [10]. Using social marketing principles that include product, place, and promotion by health professionals to change the community's behavior toward supporting and embracing the Covid-19 vaccination program can help achieve high coverage to a level of herd immunity.

Furthermore increasing the effectiveness of health education depends on the implementation of theories and models⁸. Applying the health belief model and the Diffusion of Innovation Theory can be effective in answering the above question. It is recommended that health education professionals address this issue quickly using these two or

similar models to enable behavior change in negative and neutral groups in communities. The limitation of this study is that it cannot be applied in the regions with poor or no access to social networks such as Twitter. Also, the results may be affected by the processing language.

In conclusion, we reported the findings for sentiment analysis on Twitter about vaccination of COVID-19, and built three classify "tweets" into positive, negative, and neutral sentiment. We also provided the recommendations to help the achieve high immunization coverage to a level of heard immunity.

Conflict of interests: The authors declare that there is no conflict of interest.

REFERENCES

1. Jahanbin K, Rahmanian V, Rahmanian F. Predicting COVID-19 in Iran using automate web crawling. *Pakistan Journal of Medical and Health Sciences*. 2020;14(4):987-91.
2. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau RJ, editors. Sentiment analysis of twitter data. *Proceedings of the Workshop on Language in Social Media (LSM 2011)*; 2011.
3. Kharde V, Sonawane P. Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:160106971* 2016.
4. Barbosa L, Feng J, editors. Robust sentiment detection on twitter from biased and noisy data. *Proceedings of the 23rd international conference on computational linguistics: posters*; 2010: Association for Computational Linguistics.
5. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford* 2009; 1: 2009.
6. Pak A, Paroubek P, editors. *Twitter as a corpus for sentiment analysis and opinion mining*. LREC; 2010.
7. Jahanbin K, Rahmanian V. Using twitter and web news mining to predict COVID-19 outbreak *Asian Pacific Journal of Tropical Medicine* 2020; 13: 1-3.
8. Drake JM, Chew SK, Ma S. Societal learning in epidemics: intervention effectiveness during the 2003 SARS outbreak in Singapore. *PLoS one* 2006; 1: e20.
9. Seng SL, Lim PS, Ng MY, Wong HB, Emmanuel SC. A study on SARS awareness and health-seeking behaviour - findings from a sampled population attending National Healthcare Group Polyclinics. *Annals of the Academy of Medicine, Singapore* 2004; 33: 623-9.
10. Lee NR. Reducing the spread of COVID-19: A social marketing perspective. *Social Marketing Quarterly* 2020; 26: 259-65.

Figure1: Monitoring of geographical distribution of sentiment analysis of COVID-19 vaccines of Twitter data

