Application of Twitter and web news mining in infectious disease surveillance systems and prospects for public health

Kia Jahanbin,[1] Fereshte Rahmanian,[1] Vahid Rahmanian,[*,2] and Abdolreza Sotoodeh Jahromi

Abstract

Aims: With the advancements of communication technology and growing access to social networks, these networks now play an important role in the dissemination of information and news without going through the time-consuming channels of official news networks. Analysis of social networking data is a new, interesting branch of text mining science. This study aimed to develop a text mining technique for extracting information about infectious diseases from tweets and news on social media.

Methods: A method called "Fuzzy Algorithm for Extraction, Monitoring, and Classification of Infectious Diseases" (FAEMC-ID) was developed by the use of fuzzy modeling of the Takagi-Sugeno-Kang type. In addition to the real-time classification, the method is able to update its vocabulary for new keywords and visualize the classified data on the world map to mark the high risk areas.

Results: As an example, the monitoring was performed for measles-related news items over a 183-hour period from 01/03/2019 (01:00 am) to 08/03/2019 (12:00 pm), which were related to 2,870 tweets from 2,556 users. This monitoring showed that the number of tweets posted from each region ranged from 1 to 47, with the highest number, 47 tweets, belonging to Canada. The origins of most measles-related news were in the Americas and Europe, and they were mostly from the United States and Canada.

Conclusion: The performance analysis of the developed method in comparison with other algorithms in the literature demonstrated the excellent precision of the method with a recall ratio of 88.41% and the high inter-correlation of data in each class. The proposed algorithm can also be used in the development of more effective monitoring and tracking systems for other human and even animal health hazards.

Keywords: fuzzy classification, surveillance system, Twitter, text mining, infectious disease