

Application of Twitter and Web News Mining in Monitoring and Documentation of Communicable Diseases

Kia Jahanbin, Fereshte Rahmanian, Vahid Rahmanian, Abdolreza Sotoodeh Jahromi, Mohammad Hojjat-Farsangi

Background: This study aimed to develop a method for extracting information concerning communicable diseases from Twitter and news websites. **Methods:** Using an evolving fuzzy model as the basis of work, we developed the "Fuzzy Algorithm for Extraction, Monitoring, and Classification of Communicable Diseases (FAEMC-CD)". In addition to the real-time classification of input data, the method is able to update its vocabulary to include new keywords and visualize the classified data on a map to facilitate the monitoring of disease data. **Significant findings:** In a test implementation, 1,235 malaria-related tweets posted by 1,646 users in a 34-day period from 05/06/2018 to 09/07/2018 were examined. Divided by country, the number of malaria-related tweets ranged from 1 to 166, with the highest number (166) belonging to Nigeria. The origins of malaria-related news were mostly located in Africa, South America, and India. **Conclusion:** A performance analysis of the developed method in comparison with the existing algorithms showed the high accuracy of the method with a recall ratio of 89.5% and the high intra-class correlation of its outputs. The method can also be used to monitor other diseases and for documentation and monitoring in other areas of medicine, paramedicine, and health sciences. **Key words** Web news mining; Fuzzy classification; Communicable disease; Social network; News website