International Research Journal of Engineering and Management Studies (IRJEMS)
Volume: 03 Issue: 04 | April -2019 e-ISSN: 2395-0126

HEALTH MONITORING ON SOCIAL MEDIA OVER TIME

ARUNA P
Assistant professor,
Department of Computer Science and Engineering
Sri Ramakrishna Engineering College, Coimbatore
arunapalaniappan@srec.ac.in

VIGNESH KUMAR P
Department of Computer Science and Engineering
vigneshkumar181997@gmail.com

VIKASH R
Department of computer science and Engineering
Vikashraj11497@gmail.com

SANJAY S
Department of Computer Science & Engineering
sanjaysanu2403@gmail.com

ABSTRACT

The intersection between social media and health has received considerable research attention. The proposed system analyze the twitter social media to monitor the health-related issues. The health-related tweets are collected from the twitter using R tool and twitter API. The main objective of this study is to discover the major disease and cause of death all over the world through the people tweet. The TM-EM (Expectation-Maximization) is the approach used to minimizing the prediction error on topic distributions. In this work, we are interested in using social media to monitor people's health over time. The use of tweets has several benefits including instantaneous data availability at virtually no cost. Early monitoring of health data is complementary to [post-factum studies and enables a range of applications such as measuring behavioral risk factors and triggering health campaigns. We formulate two problems: health transition detection problem and health transition prediction problem.

INDEX TERMS

Twitter API, tweets, Expectation-Maximization, R tool.

INTRODUCTION

Social media platforms have seen unprecedented worldwide growth. For example, as of June 30, 2015, Twitter has over 300 million active monthly users, 77% of whom are outside of the US. Social networks form a platform for people to share and discuss their views and opinions, and many share their health-related information both in general-purpose social media (such as Twitter, Facebook and Instagram) and in health-related social networks (communities focusing specifically on health issues, such as Daily Strength or MedHelp). Advances in automated data processing, machine learning and natural language processing (NLP) present the possibility of utilizing these massive data sources for public health monitoring and surveillance, as long as researches are able to address the methodological challenges unique to this media. Numerous studies have been published recently in this realm, including studies on pharmacovigilance, identifying smoking cessation patterns, identifying user social circles with common experiences, monitoring malpractice, and tracking infectious disease spread. The use of social media for health monitoring and surveillance indeed gas many drawbacks and difficulties, particularly if done automatically. For example, traditional NLP methods that are applied to longer texts have proven to be inadequate when applied to short texts, such as those found in Twitter. Something seemingly simple, such as searching and collecting relevant postings, has also proven to be quite challenging, given the amount of data and the diverse styles and wording used by people to refer to the topic of interest in colloquial terms inherent this type of media. The goal of this session was to attract researchers that have explored automatic methods for the collection, extraction, representation, analysis, and validation of social media data for public health surveillance and monitoring, including epidemiological and behavioral studies it serves as a unique forum to discuss novel approaches to text and data miming methods that responds to the specific requirements of social media and that can prove invaluable for public health surveillance.

LITERATURE REVIEW

In addition to summarizing the content of the session, this paper also surveys recent researches on using social media data to study public health. The survey is organized into sections describing recent progress in public health problems, computational methods, and social implications. The goal of this session was to create a single venue for cross-disciplinary researchers to present research on social media mining for public health monitoring and surveillance. The session provided a forum to share new research in a variety of important public health areas, including the detection of disease outbreaks and awareness; pharmacovigilance, including interactions with natural products and dietary supplement.

ADVANTAGES

Minimize the prediction error on ailment distributions of consecutive pre-specified periods of time.

EXISTING SYSTEM

Early monitoring oh health data is complementary to post-factum studies and enables a range of applications such as measuring behavioral risk factors and triggering health campaigns. Latent topic analysis methods such as Ailment Topic Aspect Model (ATAM), public health can now be observed on Twitter. In this work, we are interested in using social media to monitor people's health over time. The use of tweets has several benefits including express ailments in tweets. It assumes that each health-related tweet reflects ailment such as flu and allergies. Similar to a topic, an ailment indexes a word distribution.

PROPOSED SYSTEM

The proposed system is designed to identify the disease affected by people all over the world. This work is done by analysis the real time twitter data. The framework of the proposed system is shown in figure

1. In this proposed system it has four modules
   namely, Data collection
   Preprocessing
   Natural Language Processing
   Density based clustering
   Classification

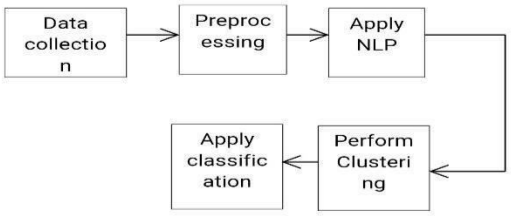

Fig 1 Proposed Framework

DATA COLLECTION

The data collection is initial process in health monitoring system. The dataset used in this research is twitter health related tweets. These tweets are collected using twitter API and R tool. After API key authentication process the twitter allows to extract the required tweets through R tool.

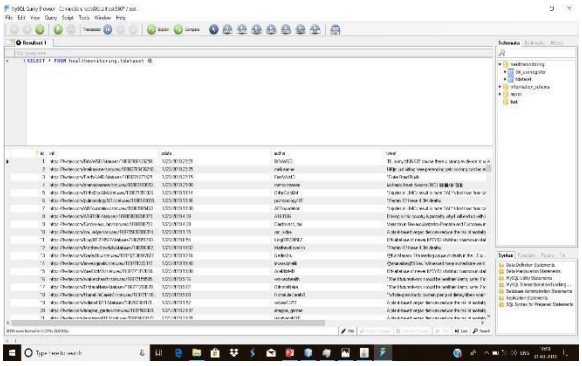

Fig 2 Dataset Collection (Tweets)

PREPROCESSING

The collected tweets contain stop words, special characters, other language tweets etc., these unwanted characters as to be removed before applying clustering algorithm. The stop words such as the, as, you, we, etc., has to be removed in order to obtain a better accuracy in classification process.

NATURAL LANGUAGE PROCESSING

Usually tweets contain many words that aren't necessary to understand the general idea of the text. These high frequency words like 'a', 'the', and 'of'

are called stop words and can be outright ignored in many situations. This is the approach fields like Information Retrieval (IR) take to reduce the dimensionality of their term spaces and improve performance. It is less useful in the field of text mining, however, because these words can often lend information and clarify semantics. Stemming (or lemmatization) is also commonly used in IR. To group similar words into one, we reduce words to their stem, or root form. For example, "walking", "walk", "walked", and "walker" will all be reduced to the root word "walk". Although it can be argued that this effect is not as harsh as stop word lists, it can still be harmful to the semantics of the text.

DENSITY BASED CLUSTERING

Density-Based Clustering algorithm has played a vital role in finding nonlinear shapes structure based on the density. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is most widely used density-based algorithm. It uses the concept of density reachability and density connectivity.

CLASSIFICATION

Classification models predict categorical class labels; and prediction models predict continuous valued functions. For example, we can build a classification model to categorize bank loan applications as either safe or risky, or a prediction model to predict the expenditures in dollars of potential customers on computer equipment given their income and occupation.

COMPARISON OF RESULTS

The fig 1.9 shows the performance evaluation of health-related disease. Where the mapping is done between the number of tweet count and the disease spread respectively. There has been twitter survey for health-related tweets for four countries. In which these four countries as received maximum number of tweet counts, whereas for other countries if the tweet count is zero, then it is not taken into consideration it remains as zero.

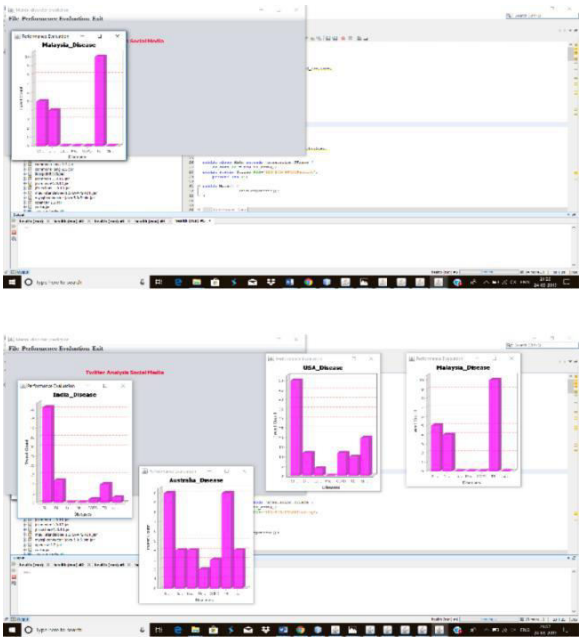

Fig 2 Performance Comparison

FUTURE ENHANCEMENT

In future, one can collect large healthcare relayed data from multiple social networking sites which may provide better results by overcoming the limitations of the project. In future, one can even collect data which includes videos and images for analyzing the effectiveness of the medicines and new treatments.

CONCLUSION

The data's collected from social media is beneficial for mankind in providing them better healthcare. The workflow of analyzing healthcare content in the social media helps to overcome the limitations of large-scale analysis and manual analysis of user generated textual content in social media. This project can provide feedback to the healthcare system organization and pharmaceutical companies for the available treatments and medicines. With the help of this project, pharmaceutical companies and healthcare providers can work on the feedback and try to come up with improved medicines and treatments. Users are provided with the resources of social media for the corresponding field of healthcare.

REFERENCE

[1] D. Rehunathan, S. Bhatti, O. Chandran, and P. Hui, "vNurse: Using virtualization on mobile phones for remote health monitoring", in Proc. 13th IEEE Intl. Conf. on e-Health Networking Applications and Services (HealthCom2011), Jun 2015, pp. 82-85.

[2] C. Khorakhun and S. N. Bhatti, "Remote Health Monitoring Using Online Social Media Systems", in Proc. 6th IFIP/IEEE Wireless and Mobile Networking Conference (WMNC2013), Apr 2016.

[3] R. K. Megalingam, V. Radhakrishnan, D. C. Jacob, D. K. M. Unnikrishnan, "Assistive technology for elders: Wireless intelligent healthcare gadget", in 2014 IEEE Global Humanitarian Technology Conference (GHTC), 2011, pp. 296-300.

[4] A. Kale, S. K. Kaul, D. P. Das. And S. Raghunath, "A smart systemfor remote monitoring of patients and sms messaging upon critical condition", in 3rd IEEE Intl. Conf. on Wireless Communication and Sensor Networks (WCSN'07), 2012, pp. 141-144.

[5] C. N. Scanaill, B. Ahearne, and G. M. Lyons, "Long-term telemonitoring of mobility trends of elderly people using sms messaging, "IEEE Trans. On Inforn. Technology in Biomedicine, vol. 10, no. 2, pp. 412-413, 2010.

[6] Z. Lv, F. Xia, G. Wu, L. Yao, and Z. Chen, "iCare: A Mobile Health Monitoring Syatem for the Elderly", in IEEE/ACM, Int'l Conf. Cyber, physical and social Computing (CPScom), Dec 2010, pp. 699-705.

[7] H. Watanabe, M. Kawarasaki, A. Sato and K, Yoshida, "Development of Wearable heart disease monitoring and alerting syatem associated with smartphone", in 14th IEEE Intl. Conf. on e-Health Networking, Applications and Services (HealthCom2012), oct 2012, pp. 292-297.