International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: 6.614
Available Online at www.journalijcar.org
Volume 8; Issue 02 (G); February 2019; Page No.17567-17570
DOI: http://dx.doi.org/10.24327/ijcar.2019.17570.3337

Research Article

EFFICIENT HEALTH MONITORING SYSTEM ON SOCIAL MEDIA FOR IDENTIFYING DISEASES

Kaushik Rohit Roshanal, Kamal Kumar K., Kushal B., and Nivedha S*

Department of CSE Sri Krishna College of Technology, Coimbatore

A R T I C L E I N F O

Article History:
Received 10th November, 2018
Received in revised form 2nd December, 2018
Accepted 26th January, 2019
Published online 28th February, 2019

Key words:

Message dissemination, fluctuation, social media network.

A B S T R A C T

Healthcare and social media together have become more of a marketing strategy than an awareness medium. Data analytics has become one of the most important aspect and techniques in the digital industry. Now a day's people all unaware about spreading of several new diseases. With the rise of social media networks (SMNs) as well as the emergence of several new diseases, social media will help us to predict various diseases from different regions based on the awareness tweets by different people in twitter. Based on the awareness tweets this system will acquire information about the disease legally and give knowledge to them about disease fluctuation by multiple message dissemination techniques. In this work, prediction about the symptoms of several diseases worldwide and it will help people educating the risks occurring through various regions of the globe.

INTRODUCTION

Social media platforms have seen unprecedented worldwide growth. As of June 30, 2015, Twitter has over 200 million active users, 78% 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 or Instagram) and in health-related social networks (communities focusing specifically on health issues, such as Daily Strength or Med Help).

Enhancement 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 researchers are able to address the methodological challenges unique to this social media. Large number of studies have been published recently in this period, including issues on health diseases, detecting smoking cessation patterns, identifying user social circles with common method, monitoring malpractice, and tracking various disease spread.The use of social media for health monitoring and surveillance indeed has 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 relatively simple, such as searching and manipulating relevant posts, has also proven to be quite difficult, 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 to this type of media.

Corresponding author:* **Nivedha S
Department of CSE Sri Krishna College of Technology, Coimbatore

public health detection and monitoring. It serves as a basic forum to discuss various approaches to text and data mining methods that replicate to the specific requirements of social media and that can prove invaluable for public health monitoring.

Related Works

Sibulela Mgudlwa and TikoIyamu [1] study can be used to guide integration of social media with healthcare big data by health facilities in the communities. The study contributes to healthcare workers' awareness on how social media can possibly be used to improve the services that they provide to the needy. Also, the study will benefit information systems and technologies and academic domains, particularly from the health services' perspective. Objectives of the study were to examine and gain a better understanding of the complexities that are associated with the use of social media and healthcare big data, through influencing factors, and to develop a framework that can be used to improve health-related services to the patients.

Chonlatee Khorakhun and Saleem N. Bhatti [2] examined the provision of alerts for a remote health monitoring (RMA) application by leveraging an online social media platform (OSMP). Using Facebook as our example platform, we find there are many facilities and features that such an OSMP can offer for RMA alerts. The use of an OSMP allows us to implement communication between actors in a career network which includes the patient and medical professionals. The OSMP also allows alerts to be delivered as Facebook notifications. We have found that the latency of delivery of such alerts is perfectly acceptable over WLAN and3G, with alerts typically delivered in a few seconds. Our proof of-concept implementation of an alert mechanism shows the

feasibility of using OSMPs for alerts.

Michael j. Paul *et al* [3] describes topics pertaining to the session, "Social Media Mining for Public Health Monitoring and Surveillance,". In addition to summarizing the content of the session, this paper also surveys recent research 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

Shaftab Ahmed ; M Yasin Akhtar Raja [4] designed a knowledge based systems; cyber domain information interchange, data management, security and ubiquitous cloud for telemedicine to enable virtual hospital services have been discussed. The wireless sensor devices for disaster management with the help of infrastructure-less communication and importance of social networking has been presented.

Health Care System

The social media intersection with health has received considerable research attention. The proposed system analyzes 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 objective of this study is to discover the major disease and cause of death all over the world through the people tweets. The TM-EM (Expectation-Maximization) is the proposed approach used to minimizing the prediction error on topic distributions.

METHODOLOGY

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.

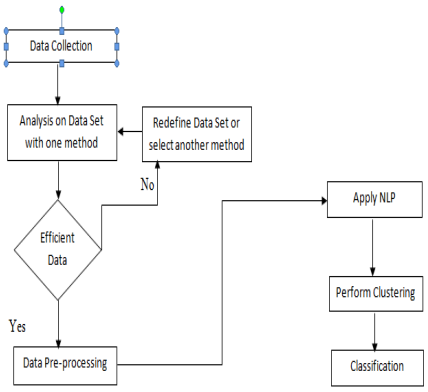

Figure 1 Health Monitoring System

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.

Preprocessing

The collected tweets contains stop words, special characters, other language tweets etc., these unwanted characters has to be removed before applying clustering algorithm. The stops 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 (NLP)

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, this is not much useful for text mining, however, because these words can often give the information and clarify semantics.

In information retrieval, stemming is commonly. For grouping similar words, we need to reduce the words to their stem, or root form. For example, "walking", "walk", "walked", and "walker" will all be reduced to the root word "walk". Although this effect is not as harsh as like stop word list, but it still be detrimental to the semantics of the text.

Mass Clustering

This clustering algorithm has played a very important role in finding nonlinear shapes structure based on the density. Mass Spatial Clustering of Applications with Noise is most widely used density based algorithm. It uses the study of density reach ability and density connectivity.

Algorithmic Steps

Let $X = \{x_1, x_2, x_3, ...,x_n\}$ be the set of data points. Mass clustering requires two parameters: ε (eps) and the minimum number of points required to form a cluster (minPts).

1. Start with an random point that has not been visited.
2. Extract the neighborhood of this point using ε (All points that are present within the distance of ε is the neighborhood).
3. 3)Clustering process starts when there are sufficient neighbors around the start point and marked this point as visited or else this point is marked as noise(soon this point also become the part of the cluster
4. If the point is found to be a part of cluster then its ε neighborhood is also the part of the cluster and the above procedure from step 2 is repeated for all ε neighborhood points. This is repeated for all points in the cluster is determined.
5. Process the new unvisited points, which lead to discover the further cluster or noise
6. This process was repeated till all the points are marked as visited.

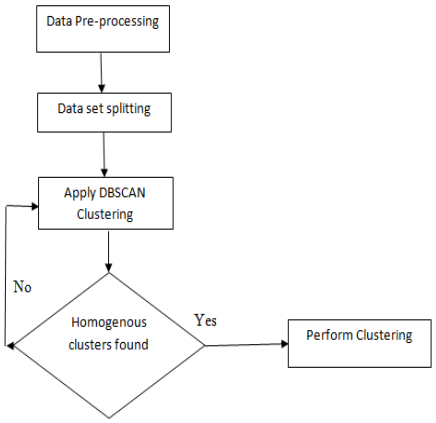

Figure 2 Flow of DBSCAN clustering

The Expectation-Maximization

The Expectation-Maximization algorithm is used in maximum possibility of estimation where the problem includes two sets of random variables in which the observable variable is X and hidden variable is Z. It works in the following ways

E-step

Estimates the expectation of missing value i.e. unlabeled class information. This step corresponds to perform classification of each unlabeled documents. Probability distribution is being calculated using current parameter.

M-step

Maximizes the likelihood of model parameter using the previously computed expectation of the missing value as if were the true ones.

Step 1

Given: X - Labeled data
Z- Missing values
θ - Unknown parameter
L(θ;X,Z) = p(X,Z|θ) - likelihood function (probability).
L(θ|X) ϵ {α p(X| θ): α >0} , p(X| θ) * p(Z|θ)

Step 2

With given variables such as X, Z and θ, the maximum likelihood estimation of the unknown parameters is calculated by marginal likelihood of the observed data. The value obtained is not traceable.

Finding maximum likelihood

$$L(\theta; X) = p(X \vee \theta) = \sum_s p(X, Z \vee \theta) \quad . (1)$$

Step 3

Calculate expected value of log likelihood function.

$$Q\left(\theta \vee \theta^{(t)}\right) = \sum_Z x\theta\hat{}(t)\,[\log L(\theta; X, Z)] \quad . (2)$$

Step 4- find the parameters that maximizes the given quantity

$$\theta\hat{}(t+1) = \operatorname{argmax} Q\left(\theta \middle| \theta\hat{}(t)\right) \quad .(3)$$

The observed data points X, may be discrete which are finite or taken from countably infinite set or it may be continuous which are taken from uncountably infinite set.

The EM algorithm can also be applied to other data model when one of the parameters Z or θ is known. This shows the iterative nature of the algorithm that continues to predict the values until converges to a particular and specific value.

The EM provides the classification output such as major disease affected by human and their country. The system also provides the result of each disease details with their region. The following section describes the experimental result

EXPERIMENTAL RESULT

The system is developed with java. The net beans IDE is used to design front end. MYSQL is used for database access. This will maintain all the data about disease fluctuation which was retrieved from tweets
Data set

The data set will contain the tweets that have been twitted by a user. The tweet will only be taken into be consider if it have been twitted with the GPS enabled system. So the main data set fields will be tweets related to disease, user location, date of the tweet being uploaded.

Editorial Policy

The submitting author is responsible for obtaining agreement of all coauthors and any consent required from sponsors before submitting a paper. It is the obligation of the authors to cite relevant prior work.

The experiment evaluation of the research is shown in the following results.

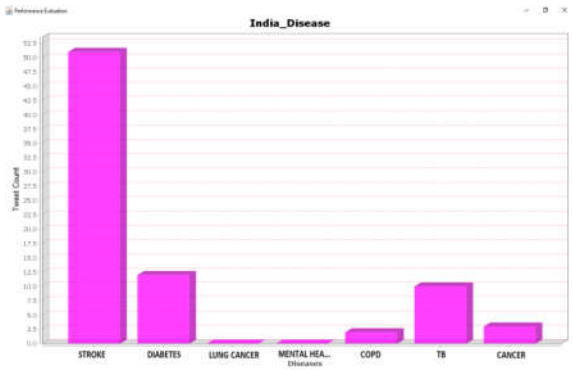

Figure 3 Disease result in India

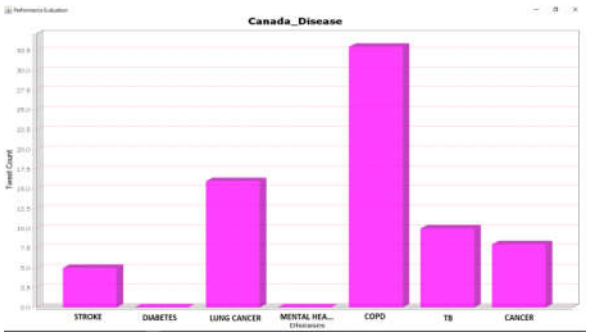

Figure 4 Disease result of canada

Performance Evaluation

When compared to the existing methodology, the proposed system is working well and it also disseminates the disease fluctuation faster and accurate. Because in existing, only the previous year data was considered for the prediction of

diseases in the particular area. Whereas the proposed system will take the recent tweets like past one year also immediate three months data also considered. So the prediction of disease becomes accurate. The disease prediction is probably accurate in proposed system than existing approach which was shown in the Figure 5.

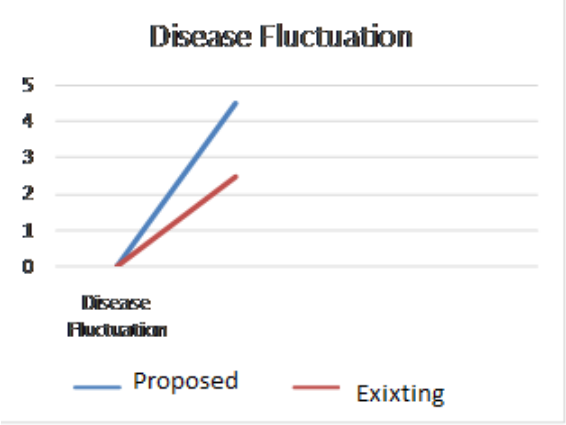

Figure 5 Disease Prediction

CONCLUSION

This proposed method is mainly to uncover ailments over time from social media. This proposed system is formulated to solve the health transition detection and prediction issues. Detection and prediction is addressed with expectation maximization system. The expectation-maximization system is used in maximum possibility of estimation to solve the health issues. This method results in significant improvement in modeling and predicting transition of health related tweets. This approach is sensible to other domains such as disaster management

References

1. Khorakhun, Chonlatee, and Saleem N. Bhatti. "Alerts for remote health monitoring using online social media platforms." In e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on, pp. 177-181. IEEE, 2013.
2. Mgudlwa, Sibulela, and TikoIyamu. "Integration of social media with healthcare big data for improved service delivery." *South African Journal of Information Management* 20, no. 1 (2018): 1-8.
3. Paul, Michael J., Abeed Sarker, John S. Brownstein, Azadeh Nikfarjam, Matthew Scotch, Karen L. Smith, and Graciela Gonzalez. "Social media mining for public health monitoring and surveillance." In Biocomputing 2016: Proceedings of the Pacific Symposium, pp. 468-479. 2016.
4. P. A. Dempster, N. M. Laird, D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", Journal of the Royal Statistical Society. Series B (methodological), vol. 39, pp. 1-38, 1977.
5. C. Ordonez, E. Omiecinski, "FREM: fast and robust EM clustering for large datasets", Proc. of the eleventh international conference on Information and knowledge management ACM, pp. 590-599, 2002.
6. R. Abbi *et al*., "Analysis of stopping criteria for the EM algorithm in the context of patient grouping according to length of stay", Intelligent Systems IS'08 4th International IEEE Conference, vol. 1, pp. 3-9, 2008.

How to cite this article:

Kaushik Rohit Roshanal *et al* (2019) 'Efficient Health Monitoring System on Social Media for Identifying Diseases', *International Journal of Current Advanced Research,* 08(02), pp.17567-17570.
DOI: http://dx.doi.org/10.24327/ijcar.2019.17570.3337

****