Sentiment analysis on Healthcare Tweets

P.Kamakshi[1]

[1]*Professor, Department of Information Technology, Kakatiya Institute of Technology and Science, Warangal, India*

Abstract

In today's world variety of posting on social media offer huge information about the health issues, remedies, food and medication. Twitter is an online social networking service in which users can post their opinions about various topics and also cooperate with each other with messages known as "tweets". Tweets are very helpful in sharing the health related issues, medicines, hospitals information. Twitter helps the people to know about the disease symptoms, services and details about the hospital before they go for consultation. *Sentiment* analysis is a metric commonly used to investigate the positive or negative opinion within these *messages*. *Sentiment* analysis methods can be used in *Twitter health* care research. The analysis will help the users to better understand the alternative available. Sentiment analysis can also facilitate the healthcare industry to use reliable data for their growth by taking necessary measures. Sentiment analysis applies software to analyze the patient's tweets regarding their healthcare experiences regarding medicine, doctor, hospitals..It helps users as well as many healthcare organizations to understand their customers opinion and to take necessary measure to rectify the gaps. Ultimately, as more attention is given to such opinion analysis the health standard in the society will improve. Main aim of this paper is to build an algorithm that can accurately classify Twitter messages as positive or negative.

Keywords : *Tweets, sentiment analysis Health care, Healthcare organiations*

Introduction

With more concern towards individual health, today everyone is using Internet to participate in medical forums to gather health-related information, to share experiences about drugs, treatments, diagnosis or to interact with other users with similar condition in communities. Monitoring social media platforms has recently fascinated medical natural language processing researchers to detect various medical abnormalities such as adverse drug reaction. In this paper, we present a benchmark setup for analyzing the sentiment with respect to users' medical condition considering the information, available in social medium called twitter. Twitter is an online social networking medium, where registered users share or post messages known as tweets. There are several unstructured, free-text tweets related to health care being shared on Twitter, which is becoming a popular area for health care research In this work we used this data to observe general users health Awareness towards sentiment analysis has been promising over the last few decades due to the huge popularity of social media. The extraordinary rise in sharing the information in social media is observed in health conscious people such as medical debates which are flooded by many users. Many of such users are likely to be patients who seek help for health-related information, want to share medical problems [4] and their experiences and also want to opt for informational support or opinions from the other users like patients, health-practitioners. These texts present a platform to glance into a user's opinions, sentiments and feelings in a very wide range.Sentiment is a metric normally used to explore the positive or negative judgment within these messages. Exploring the methods used for sentiment analysis in Twitter health care research may allow us to better understand the available options.

Sentiment Analysis estimate whether a part of text called tweet[2] is representing positive, negative or neutral. It is called as opinion mining as it derives the opinion or mind-set of a user who posted the tweet. The objective of this paper is to build an algorithm that can accurately classify Twitter messages as positive or negative, with respect to a query term. Our hypothesis is that we can

obtain high accuracy on classifying sentiment in Twitter messages using machine learning techniques.

Background

With advancement of Internet , its scope is becoming wider day by day. Social Media platforms like Twitter Facebook, plays a major role in disseminating the trendy topics , news at very fast speed. People give their opinion[1] and judgments in large and make the topic trendy. These topics in general are meant to bring consciousness or to promote movies, elections, celebrities. Many organizations take benefit from the people's feedback. They try to enhance their goods, services and also in improving promotion policies. Thus, there is a huge possibility to improve business driven applications by identifying and analyzing fascinating patterns from the huge data available in social media. In particular, the hypothesis try to categorize all the exchanged chat into the label of positive, negative or neutral. This sort of information develop a source for people to assess, rate about a particular movie, products etc.

Sentiment Analysis on Tweets

Twitter is a social networking service where users can send and read short 140-character messages called "tweets." One of the popular areas i.e. in healthcare several unstructured, free-text tweets relating to health care are shared on Twitter. It is becoming a popular area for health care research. Sentiment is a metric[6] commonly used to investigate the positive or negative opinion within these messages. Exploring the methods used for sentiment analysis in Twitter health care research may allow us to better understand the options available for future research in this growing field. Patients and citizens in general, are increasingly using the Internet for searching health information and support. Eurobarometer, which is a series of public opinion surveys reports that nearly 60% of European people had a glimpse on online health-related information and more 90% of them , reported that they will prefer to continue the use of Internet as a main resource in the future to access health-related information. It is found that nearly 80% of searches through search engine are related to health topics, symptoms, treatments. Some of the users show interest in reading others comments or experiences in Health communities and some of the users take online consultation for disease, drug or treatment. The use of online health communities is particularly popular among chronic patients. Surveys show that these patients significantly benefit from social interaction with peers and the sharing of knowledge, experiences and support. Information in online health forums and communities is also of great interest for researchers and professionals, as it allows for research in a very normal process and cannot be simulated in laboratory atmosphere. Some of the examples are the side effects of medicines, alternative treatments are some of the examples.

The health industry is one of the important stakeholders which monitor the people and patients[5] opinion regarding their products and services and the level of their satisfaction. However, the quantity of information is so gigantic that it is difficult for the users to find the information that is really needed. Opinion *mining* is the process which is used to classify a portion of text into positive, neutral or negative. Several unsupervised and supervised learning methods are[7] anticipated to achieve better result.

Process

The dataset about health care sentiment is expected to be from the peoples own opinion about the discussed matter. As the data collected form online resources don't follow any standard format. Such information or reviews are of little use to sentiment analysis and are therefore filtered off, in order to focus on the opinion rich content only. Subjective reviews may also contain some objective statements representing facts and figures which are filtered for the same purpose. Review documents produced by ordinary authors may contain all kind of inconsistencies like grammatical errors, spelling mistakes, over or under use of capitalization, word shortening, The order of tweets are also very important in such discussion as they lead towards common grounds. It not only express sentiment for healthcare problems but also support the opinion with strong reasons which make useful information to public.

Sentiment Analysis is the process of 'computationally' determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining, deriving the opinion or attitude of a speaker. In this module the sentiment analysis is done on the data i.e twitter tweets which is collected from the input given by the user and it will able to identify whether the given sentence is negative or positive by using natural language tool kit and able to produce whether it is positive or negative. The NLTK module is a massive tool kit, which performs everything from splitting sentences

from paragraphs, splitting up words, recognizing the part of speech of those words, highlighting the main subjects, and then even with helping user machine to understand what the text is all about. The live data coming in from the Twitter streaming API, can be represented by live graph that shows the sentiment trend. The Matplotlib. animation function helps in creating the live graph.

Result

The tweets in the form of texts are classified and sentiment analysis is done .The classified tweets in the form of negative and positive are shown on the screen by using the naïve bayes algorithm and the sentiment analysis modules in which the process of tokenizing, lemmatizing, chunking, tokenizing are done. The input given in the text formats i.e. either an word or sentence. After the sentiment analysis is done on the data which was generated by the twitter related to the input the data will be represented in the graphical form using matplotlib. The graph helps to know the accuracy percentages of individual classifiers[3] which helps in identifying the classifier which gives optimized solution.In the graph by matplotlib, the representation of the analysis of the tweets is in the form of the graph model. The X-axis represents the number of tweets which are generated directly and Y-axis represents the value of the sentiment analysis which is generated for each tweet .

After getting the output of the graph we checked the accuracy percentages of the Naïve Bayes Classifier and the other classifiers for 3000 tweets for the term cancer. The best classifier for 3000 tweets is SGDC Classifier with the accuracy percentage of 89.0.

Conclusion

One of the most popular application of Sentiment analysis which will be helpful for the user in better healthcare is to apply it on the information collected from social media. The workflow of analyzing healthcare content in the social media helps to overcome the limitations of large scale data analysis and manual analysis of user generated textual content in social media. This work can help the users to be updated with the effectiveness of the medicines and it can even suggest them with few better medications available. 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 the improvised medicines and treatments. Users are provided with the resources of social media for the corresponding field of healthcare.

Conflict of Interest: Nil

Source of Funding: Self

Ethical Clearance: Not required

References

1. Doaa Mohey El-Din Mohamed Hussein. A survey on sentiment analysis challenges. Journal of King Saud University – Engineering Sciences. 2015Nov: 330-338.
2. Kumar P. K., Nandagopalan S.Insights to Problems, Research Trend and Progress in Techniques of Sentiment Analysis. International Journal of Electrical and Computer Engineering (IJECE) . 2017 Oct;7(5) :2818-2822.
3. Christiyana Arulselvi, A Sendhilkumar ,S Mahalakshmi S. Classification of Tweets for sentiment and Trend Analysis.International Conference on Intelligent Computing and Control Systems ICICCS .2017:566-573.
4. Shweta Yadav, Asif Ekbal, Sriparna Saha, Pushpak Bhattacharyya. Medical Sentiment Analysis using Social Media: Towards building a Patient Assisted System. In: LREC 2018.Proceedings of the Eleventh International Conference on Language Resources and Evaluation ;2018; Miyazaki, Japan: 2790-2797
5. Ranjitha Kashyap, Ani Nahapetian. Tweet Analysis for User Health Monitoring.In: International Conference on Wireless Mobile Communication and Healthcare;2014; MOBIHEALTH 2014, November 03-05, Athens, Greece:348-355.
6. Anna Jurek, Yaxin Bi, Maurice Mulvenna. Twitter Sentiment Analysis for Security-Related Information Gathering.In: IEEE Joint Intelligence and Security Informatics Conference:2014; The Hague, The Netherlands:48-55.
7. Shubham, Shashank Kumar, Sunanda Dixit. Classification of tweets into various categories using classification methods. International Journal of Advance Research, Ideas and Innovations in Technology.2018;4(3): 937-940.