ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Determining Drug Efficacy by Extracting Opinions from Social Network Data Using Mahout.

Dipali Govardhan Akhare* and Saravanan R.

School of Information Technology, Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India.

ABSTRACT

Social media has become a boon in the recent times due to its reach and ease of access. The millions of tweets received every year could be subjected to sentiment analysis. The current Analytics tools and models used that are available in the market are not sufficient to manage Bigdata. Therefore, there is a need to use a Cloud storage for such type of applications. So we have utilized Hadoop for intelligent analysis and storage of big data. These tweets also contains drug related information and their adverse reactions. In the proposed system we have used this information as source and predicted a best drug for related disease. User give their opinion related to drug. These opinions can be analyzed and used to predict a better drug for associated disease. Disease name is taken as an input. Tweets containing the keyword are fetched with the users latitude and longitude. These latitude and longitudes are then used to show user the location of nearby hospital and map is also shown. Preprocessing and SVM classification is applied on the data followed by polarity prediction which results in recommendation of the best drug.

Keywords: Google map, polarity prediction, Twitter, Hadoop

**Corresponding author*

ISSN: 0975-8585

INTRODUCTION

We are living in the age of data, where data comes in all types, structured or unstructured and from all types of sources. There is total 4.4 zettabytes of data according to the IDC in 2014, which is now increased to 44 zettabytes. Large amount of data in various field like Logistics, Financial, Health data, Social media, Scientific data etc is analyzed, and its challenging to manage and analyze this data. To analyze such enormous information we make use of apache Hadoop. Hadoop helps to perform operation in a distributed way and in efficient manner. There are various social networking websites like Facebook, Twitter, social stage science wibe, the information present on these sites can be helpful in many ways. Twitter is most famous among them, every second, six thousand tweets are tweeted. All these tweets are publically visible by default, but one can restrict this by allowing their followers only. 60% of doctors says that social media has changed the way of living. Twitters data can be used for public awareness of drugs and disease. Users can share their experience related to the drug or disease.

The reason we choose Twitter, because with help of Twitter4j library we can fetch the tweets from the Twitter. In present scenario, people are going to hospital waiting in a queue for appointment for consultation. This process is very time consuming and sometimes patient may forget to say few symptoms of the disease or the drug. Because patients will have only few minutes inside the consultation room meeting with doctor, in that tension and by listening to doctor's points and suggestions they may forget few points. It is not necessary that all the doctors will have up to date knowledge about new medicinal drugs, in that case they will be providing old medicines to the patients which may take long time to cure.

Even for existing diseases many new symptoms are rising due to the change in weather condition and new food style. But in the proposed system, recent medical information is being collected from Twitter and make use of it effectively for the public. In Twitter, the users will post the tweets based on their experience they faced due to the diseases and used drugs. The recently posted drugs related tweets or disease related tweets by users are being extracted from Twitter by using API and application details. Through these tweets, people will get new information about new diseases spreading around the world and also the new drugs. If Twitter users posted the tweets by specifying the symptoms then the people can know the diseases and drugs along with its symptoms. In hospitals, some doctors may cause medical error which make the patients to stay long for treatment and it is expensive. Whereas in the proposed system, medical error can be avoided since this system will give out the best drugs based on the majority of public's opinion. Using polarity prediction the people will get best solution. The proposed system is inexpensive, less time consumption and provide best solution.

Scope of the Project

In today's world, all the people are very addicted to social networking. All the information from the social networking websites can be used effectively. This system will analyzed the tweets information and gives out the best medicine for the consumer.

Problem Definition

In the existing system of this paper a system using social network information posted by the users. Physicians, pharmacists and consumers post their drug usage experience in the web forum. In the existing system, they use only partially supervised classification which is not much efficient, so in the proposed system Support Vector Classification is used. As they used partially supervised classification, the data will be in high dimension so to reduce the high dimension they used Latent Dirichlet Allocation method. From the forum messages they can only identify whether the drug is adverse or not. If it is adverse it gives a warning to Food and Drug Administration.

In the proposed system First user will enter his/her login credentials, for accessing the application. If the user is not registered then he/she will register first after that user will enter keyword as any disease name then this disease name is searched into the tweets. For online extracting the tweets twitter4j API is used. I have extended the limit to 100 for fetching the tweets from Twitter. After this process all data preprocessing is done and SVM classification is applied and result is obtained. The reminder of this paper is organized as follows. In Section 2, the literature survey is discussed. In Section 3, the architecture of paper is explained. In

Section 4, implementation and the experimental set up of the proposed system. Finally section 5, continues the result analysis followed by the conclusion and reference.

LITERATURE SURVEY

Ajay Deshwal et al. [1] proposed an architecture which considered unigrams, punctuations, emoticons and opinion lexicon altogether as features on sanders dataset. This paper compared the results of six supervised classification algorithms. The simulation results revealed that discriminative multinomial Naïve Bayes algorithm and sequential minimal optimization algorithm (SMO) are most balancing and well-performing algorithms. Simulation results show that combining many features like unigrams, emoticons and sentiment gazetteers altogether for sentiment classification analysis yields improved overall results.

Mitali Desai et al. [2] proposed a system where first the data is crawled from the Twitter, followed by the data collection phase. Then data is preprocessed and then feature extraction after a training set is built and classifier are applied. These classifier are machine learning algorithm after that polarity data is classified. The main survey contains the classifier algorithm. The comparison is done between all the classifiers, like SVM Navie Bayes maximum entropy random forest from this result is obtained, the result shows that SVM provides better accuracy as compare to other.

Monu Kumar et al. [3] proposed an architecture of Twitter sentiment analysis of Airtel telecommunication company. In this paper in the beginning tweets related to Airtel company are collected, these tweets are retrieved using a hashtag and a company name. For this Twitter4j API is used. Then data is integrated in the Hadoop and converted intoHadoop sequence file. This file is uploaded to HDFS. Nest to this process, stopword removal and preprocessing is done. This data is then analyzed and sent to the classifier. In this system Navie bayes is used. After the analysis part result are displayed i.e. what people has commented or their opinions about the company is shown in negative, positive and neutral polarity scores.

Yang Peng et al. [4] proposed, in step 1 first real time tweets are fetched and all these tweets are raw data. In the second step this raw data is preprocessed and given to the next step. In the third step three classification algorithms are used, RULE NLP, combined machine learning method, drug related classification and user existence classification. In step 4 this output is then given for sentiment analysis, in this step there are two methods subjective classification and polarity classification. Step 5 is drug event extraction and the output of step 3 and 4 is used. The main objective in this paper is to compare two pipelines, benchmark pipeline and the proposed pipeline. In benchmark pipeline in step 3 it uses three classification algorithms and in proposed pipeline only one algorithm is used. The main advantage of proposed pipeline is that it extracts more ADE. Refer table 1 for result comparison.

Table 1. ADEs Extraction Result Comparison

	Proposed Pipeline	Benchmark Pipeline by Wu, Moh, and Khuri [3]	Ratio (%)
Pregabalin			
Total ADEs	376	49	767%
New ADEs	71	4	1775%
Baclofen			
Total ADEs	198	12	1650%
New ADEs	34	3	1133%
Duloxetin			
Total ADEs	327	45	726%
New ADEs	65	4	1625%
Gabapentin			
Total ADEs	302	130	230%
New ADEs	96	10	960%
Glatiramer			
Total ADEs	36	0	Infinity
New ADEs	3	0	Infinity
Overall Total			
Total ADEs	**1239**	**236**	**525%**
New ADEs	**269**	**21**	**1281%**

Dipali et al. [5] proposed some data mining techniques for detecting adverse drug event. These techniques are association rulemining, association technique using fuzzy rules, clustering, and statistical methods are prp, chisquare and risk ratio. Last one is miscellaneous technique. The association rulemining technique is basically used for detecting adverse drug event between the drugs and disease. In this method uninformative association are removed and only causative are kept, frequently closed item set mining is applied to reduce the price of generation of big dataset. In association techniques using fuzzy rules, fuzzy rules are used for determining the casual association. Stored values are very important for determining the cue values. Cue values are compared for the new patient and the one which are stored. This comparison gives the strength of the association rulemining. In the second technique smart software is used for determining the ADE of the patient. This software compares the drugs and the adverse drug event related to that particular drug . If it is not related then it collects the data related to that drug. PRP method forms a matrix between the drug and the ADE and this matrix is also called as contingency matrix.

Vijay Shankar Gupta et al. [6] have done sentiment analysis (opinion extraction) of hospital. For this the author has used Hadoop framework and Twitter data. Tweets are collected for all hospitals from the twitter using twitter API. They are cleaned and their semantics are analyzed. Sentiment analysis for all the hospitals is conducted on the basis of emotion and polarity. Following are the steps: 1) Twitter4J API is used for crawling the tweets and this API is of two types, stream API and search API. In this paper author has used search API. Hospitals name is used as a keyword for searching the tweets. Flume is used for fetching and it is stored into HDFS. These tweets are converted into key value pairs and a table is prepared in hive. This date is then saved into wordcount.csv file.

Once this file is created then R is used for data preprocessing process. The main purpose of this paper is to do sentiment analysis i.e. to find out how manytweets are positive and negative. The author has done survey on hospitals like AIMS, APOLLO, MEDANTA, MAX and FORTIS. From the result it is concluded that aims has max no of positive tweets and Apollo has negative tweets. While rest of the hospitals have moderate no of tweets. AIMS is considered as no 1 hospital. The limitations for this system are, 1) we can fetch a limited no tweets at a particular time, 2) it may also happen that the no is less than expected, 3) it is not necessary that data should always be up to date.

L.Jaba Sheela et al. [7] have done a review for opinion extraction of Twitter data using two APIS stream API and rest API, rest API is used for short connections. Preprocessing, opinion extraction analysis is done and visualization is done using heat map tool and result is calculated using formulas like recall f measure and accuracy.

The steps in the proposed system are as follows: Collect unstructured data from social media, preprocessing of this data with sentiment analysis. And this preprocessed data is stored in NoSQL database, extract the sentiments from the NoSQL database and visualize this data. The details of the modules are presented below.

Data Streaming: Extracting real time tweets using Twitter Streaming API For classification and training the classifier we need Twitter data. For this purpose wemake use of API's Twitter provides. Twitter provides two API's; Stream API1 and REST API2. The difference between Streaming API and REST APIs are: Streaming API supports long-lived connection and provides data in almost real-time. The REST APIs support short-lived connections and are rate-limited (one can download a certain amountof data [*150 tweets per hour] but not more per day. 2)Preprocessing: In this phase, the tweets are available as text data and each line contains a tweet.Initially we clean up or remove re-tweets as that will induce a bias in the classification process. We need to remove the punctuations and other symbols that doesn't make anysense as it may result in inefficiencies and may affect the accuracy of the overall process. 3)Sentiment Polarity Analysis: MapReduce is a new parallel programming model, hence the classical Naive Bayes based sentiment analysis algorithm is adjusted to fit into Map Reduce model. We choose to employ a Naive Bayes classifier and empower it with an English lexical dictionary SentiWordNet. 4)Visualization: Tweets are presented using several different visualization techniques. Each technique is designed to highlight different aspects of the tweets and their sentiment. These techniques are Heatmap, Tagcloud, Timeline, map affinity. We will evaluate our experiment results by using following Information Retrieval matrices

- Precision = TP/(TP + FP)
- Recall = TP/(TP + FN)

ISSN: 0975-8585

• F-measure = 2*Precision*recall/(Precision + recall)
• Accuracy = TP + TN /(TP + TN + FP + FN)

Satya Katragadda et al. [10] proposed the approach 1) To determine the performance of base line classifier for ADE . 2) To determine the latency between the drugs and the symptoms.

The tweets corpus consist of total 1lakh seventy two thousand and Eight hundred tweets from 864 users which were tweet during a time span of one year. And from these tweets total 103 drugs and 237 symptoms related to that drug were extracted and 2436 drug symptoms interaction was retrieved. Overall 606 drug symptom interaction were adverse. The performance of this model is calculated by recall precession F-score and accuracy. Base line classifier is used to compare the performance of the model. In order to compare the performance 10-fold cross validation is performed on the dataset. The proposed ADE performed well as compared to the base line model and this shown on the table 2, the value for n is kept 5 i.e. the drug and symptom is mentioned within the 5 days. The false negative and false positive values were higher in case of baseline model. The proposed model of ADE is tested using different values of n that is described in graph generation stage of the model. The value of n is considered if a drug symptom interaction needs to be considered in the medical history. Table 3 shows this with varying values of n with actual no of ADE. Table 4 shows the classification result obtained from 10-fold cross validation method with varying no of n values.

Table 2. Results of 10-fold cross-Validation For different Models

Model	Precision	Recall	F-Score	Accuracy
Preferential Attachment	0.461	0.517	0.488	0.755
Adamic-Adar Index	0.573	0.467	0.514	0.782
Jaccard-Coefficient	0.538	0.561	0.571	0.803
Common Neighbors	0.603	0.552	0.577	0.808
Temporal Difference	0.653	0.531	0.582	0.811
Edge Weight	0.61	0.566	0.583	0.823
Proposed Model	0.787	0.763	0.772	0.886

Table 3. ADE and normal drug-symptom interaction for different values of n

Value of n	0	5	10	30	60
Total number of edges	614	2436	4927	10911	19420
ADE interactions	163	606	1181	1884	2749
Normal interactions	451	1830	3746	9027	16671

Table 4. Results of 10-Fold for different values of n

N	Precision	Recall	F-Score	Accuracy
0	0.632	0.763	0.691	0.85
5	0.787	0.763	0.775	0.886
10	0.54	0.761	0.632	0.849
30	0.531	0.655	0.586	0.871
60	0.453	0.573	0.506	0.875

SYSTEM ARCHITECTURE

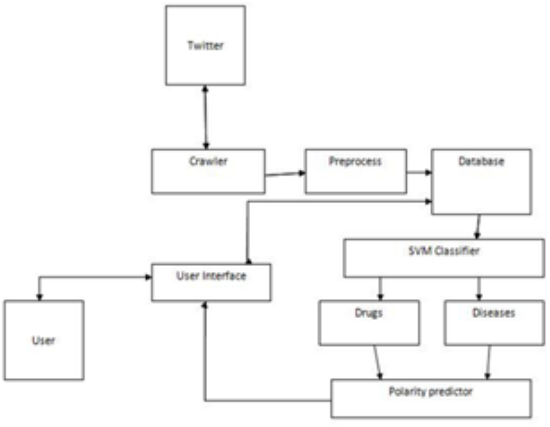

Figure 1. System Architecture of proposed system

The system architecture works as follows: first user has to register to the application, the register phase consist of various fields like first name username and password mobile no. After entering these details user will be logged in to the application. All these details are saved in the server. If the user enters wrong details then the user won't be able to access the application. If user wants to know about the service then a service button is available. 'About us' button is given for the extra information. After login phase user can enter any disease name as keyword with hash tag. The recent 100 tweets containing the same keyword will be fetched, with their location latitude and longitude. For this purpose the user who is tweeting has to keep his GPS button on. If the GPS button is off then user won't get the co-ordinates of the location. These latitude and longitude are helpful for finding the hospitals in that location. A radius is set for area to fetch the hospital related information inthat area. User will get to know all the medical facilities located nearby. To use this feature the user has to press Getmap button. For fetching the tweets from Twitter, Twitter4j API is used, Twitter4j is purely java based library which is used to fetch the tweets.

Data pre-processing is very tedious task to do, tweet contains many word alphabets and symbols with hash keyword or emoticons, stop words, repetitive words etc. We have to remove these words, feature extraction task is also there in which adjective nouns are checked. This task of preprocessing is done by map reduce mechanism in Hadoop which will be explained in implementation part.Afterpreprocessing three categories are there, drug based tweet, disease based tweet, level of the disease. The classification of drug based and disease based is done by SVM classifier algorithm. Support vector machine algorithm is a supervised machine learning algorithm, which is best for accuracy. SVM will form a boundary between two classes to classify the tweets as drug based and disease based. After clicking on drug based tweets, drug related tweets are fetched on click of disease based tweets, disease related tweets are fetched. Level of the disease will be known on click of level, if user wants to know about the life threatening disease then directly a drug is not recommended in this case level wise treatment is given, for ex cancer, tuberculosis etc. Polarity of the disease is known after clicking on polarity button, there are three polarities positive, negative and neutral score. Bar chart will shown the result and analysis.

EXPERIMENTAL SETUP

In the beginning user have to register first for login purpose, username password mobile No etc is entered and registered successfully. If username or password is incorrect than invalid username and password error is given. After successful login a disease name is entered as a keyword for searching the tweets. Connection is established to the Twitter using Twiiter4j. To extract the tweets, first the connection should be established with twitter account using the twitter API called Twitter4j. Then create the twitter developer application in twitter developer site. From the developed application we get the consumer key, secret key, Access token and token secret key. Using these keys and tokens, it is Configured and connected with twitter. In this API it contains many parameters to extract and read from the Twitter Factory by using query search and have to manage the query search concludes in Query Result. Using get Tweets method we can get the tweets, from which we can extract the tweet username.

These keys will be used for further connection. Once this process is done we can fetch the drug related tweets and location of that tweets. For this purpose user have to go the setting and on the GPS location. Tweets with location are fetched now we can see the hospitals in 1500 meter radius. Get map will display the map of the hospitals the hospital. Twitter is used by all users from all around the world so hospitals are displayed from that location.

Google map API is used for displaying the map, now these tweet are given to preprocessing purpose and stop word removal. Tweet contains symbols smileys. This file is saved to HDFS in csv format and map reduce code is applied on that and tweet are converted in to key value pairs and all unnecessary words are removed. This output is given to SVM classification algorithm. This algorithm will classify it into disease and drug based tweets. Level of the disease is also displayed for selected life endangering disease. Level wise medication is shown a link is provided this link will redirect to information related to that disease.

After SVM classification tweets refined will be displayed then polarity detection is applied it will calculate the polarity score and drug is recommended at last. This classification is done using mahout we have to install Hadoop in the system in this project we have used Ubuntu operating system and all setups are done for installing the mahout in the Operating System.

ISSN: 0975-8585

RESULTS AND DISCUSSION

In the existing system patients go to hospital for their treatment, sometimes it takes long time to get the appointment of the doctor. And after meeting the doctor the patient won't get time to understand the disease the patient is suffering and details about the drug. While in proposed system user can gain the information about the disease and the drug. Telemedicine concept is also used for remote location areas where medical help is not possible. This technique is not helpful for all the time because it requires continuous internet connection. While in the proposed system it works as follows. First user will crawl the tweets according to disease the pre-processing is performed on that. In this process words like r is replaced to are, u is replaced to you, stopwords like the, at, or, but are removed repetitive words like hhhhapppy are converted to happy, then emoticons and hash symbols are removed and we get a preprocessed tweet. We can also determine the location the hospital in users tweet location, and with help of Getmap function map is shown. Now user can know where these treatments of the disease is possible. The best hospital having multi specialist doctors are known to the user. User can determine level wise treatment of the disease like in the first stage which treatment is required and in the second stage which treatment is required. In this way proposed system is less time consuming and effective.

CONCLUSION AND FUTURE WORK

Collecting opinions of the user related to drug and disease is a complex task. As the data is highly unstructured, the name of the drugs are hard to spell and write. To collect this data from Twitter and find the adverse reaction of that drug manually is a tedious task to do. Generally tweets are only 140 characters of length i.e. short in size. To address this issue we have proposed a architecture to fetch real time tweets and latitude & longitude of that location. And perform data cleaning and pre-processing on the tweets. Apache Hadoop framework is used for this purpose and mahout is used for supportvector machine algorithm. Level of the adverse disease will be known and polarity prediction is done to obtain the final score. The proposed method can be applied to other fields like movie rating, product rating.

The future work may include various machine learning algorithms to be used to for classification purpose. This method can also be useful for other fields, for example movie review or any product related information. As Twitter uses only English language we cannot use other language data. In future we can go for other languages .

REFERENCES

[1] Ajay Deshwal, Sudhir Kumar Sharma, "Twitter Sentiment Analysis using Various Classification Algorithms", In the proceedings of 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, Uttar Pradesh, India, pp.251-257, 2016.

[2] Mitali Desai, Mayuri A. Mehta, "Techniques For Sentiment Analysis Of Twitter Data : A Comprehensive Survey", In the proceedings of International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, Uttar Pradesh, India, pp.148-154, 2016.

[3] Monu Kumar, Dr. Anju Bala, "Analyzing Twitter Sentiment Through Big Data", In the proceedings of 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp.2628-2631, 2016.

[4] Yang Peng, Melody Moh, Teng-Sheng Moh, "Efficient Adverse Drug Event Extraction Using Twitter Sentiment Analysis", In the proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, pp.1011-1018, 2016.

[5] Dipali, Yogita, "Data Mining as a Tool For Detecting Adverse Effects Of Drugs", In the proceedings of International Conference on Computing, Communication and Automation (ICCCA2016), Greater Noida, UP, India, pp.188-192, 2016.

[6] Vijay Shankar Gupta, Shruti Kohli, "Twitter Sentiment Analysis in Healthcare Using Hadoop And R", In the proceedings of International Conference on Computing for Sustainable Global Development(INDIACom), New Delhi, India, pp.3766-3772, 2016.

[7] L.Jaba Sheela, "A Review of Sentiment Analysis in Twitter Data Using Hadoop", International Journal of Database Theory and Application, Vol.9, No.1:77-86, 2016.

[8] Peiman Barnaghi and John G. Breslin, Parsa Ghaffari, "Opinion Mining and Sentiment Polarity on Twitter and Correlation Between Events and Sentiment", IEEE Second International Conference on Big Data Computing Service and Applications, Oxford, United Kingdom , pp.52-57, 2016.

[9] Jos´e Antonio Iglesias, Aaron Garc´ıa-Cuerva, Agapito Ledezma and Araceli Sanchis, "Social Network Analysis: Evolving Twitter Mining", IEEE International Conference on Systems, Man, and Cybernetics, Budapest, Hungary, pp.1809-1814, 2016.

[10] Satya Katragadda, Harika Karnati, Murali Pusala, Vijay Raghavan, Ryan Benton, "Detecting Adverse Drug Effects Using Link Classification on Twitter Data", IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Washington DC, USA, pp.675-679, 2015.

[11] Satwik Sabharwal, Samridhi Gupta, Thirunavukkarasu. K, "Insight Of Big Data Analytics In Healthcare Industry", In the proceedings of International Conference on Computing, Communication and Automation (ICCCA2016), Greater Noida, UP, India, pp.95-100, 2016.

[12] Rucha Kulkarni, Sayali Dhanawade, Shraddha Raut, Prof D.S.Lavhakare, "Twitter Stream Analysis for Traffic Detection in Real Time", International Journal of Advance Research, Ideas And Innovations In Technology, Volume 2, Issue 5: 1-4, 2016.

[13] Fan Yu, Melody Moh, Teng-Sheng Moh, "Towards Extracting Drug-Effect Relation From Twitter: A Supervised Learning Approach", In the proceedings of IEEE 2nd International Conference on Big Data Security on Cloud, IEEE International Conference on High Performance and Smart Computing, IEEE International Conference on Intelligent Data and Security, N ew York, USA, pp.339-344, 2016.

[14] Maria Karanasou, Anneta Ampla, Christos Doulkeridis and Maria Halkidi, "Scalable and Real-time Sentiment Analysis of Twitter Data", In the proceedings of IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Catalonia, Spain, pp.944-951, 2016.

[15] Nicholas Chamansingh and Patrick Hosein, "Efficient Sentiment Classification of Twitter Feeds", In the proceedings of IEEE International Conference on Knowledge Engineering and Applications (ICKEA), Singapore, pp.78-82, 2016.

[16] Karen Oconnor, Pranoti Pimpalkhute, Azadeh Nikfarjam, Rachel Ginn, Karen L Smith, Graciela Gonzalez "Pharmacovigilance on Twitter? Mining Tweets for adverse drug reactions", In the annual proceedings of AMIA Symposium, Arizona, USA, pp.924-933, 2016

[17] Abeed Sarker , Rachel Ginn , Azadeh Nikfarjam , Karen O'Connor , Karen Smith, Swetha Jayaraman, Tejaswi Upadhaya, Graciela Gonzalez, "Utilizing social media data for pharmacovigilance: A review", Journal of Biomedical Informatics, Volume 54, Issue April 2016:202-212, 2016.

[18] Xiao Liu, Hsinchun Chen, "Identifying Adverse Drug Events from Health Social Media Using Distant Supervision", Journal of Biomedical Informatics, Volume 54 (2016):202–212, 2016.

[19] Martin Sarnovsy, Peter Butka, Andrea Huzvarova, "Twitter data analysis and visualizations usingR language on top of the Hadoop platform", In the proceedings of IEEE 15th International Symposium on Applied Machine Intelligence and Informatics, Herl'any, Slovakia, pp.327-332, 2017.

[20] Bilal Saeed Raja, Ahmad Ali, Atif Parvaiz Malik, Mansoor Ahmed, Abid Khan, "Semantics Enabled Role Based Sentiment Analysis for Drug Abuse on Social Media: A Framework", In the proceedings of EEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, pp.2016-211,2016.

[21] Liang Wu, Teng-Sheng Moh, Natalia Khuri, "Twitter Opinion Mining for Adverse Drug Reactions", EEE International Conference on Big Data (Big Data), Santa Clara, CA, USA , pp.1570-1574, 2015.