Drug Reviews Using Data Mining Model

Abhimanyu D. Sangale, Prof. P.N.Kalavadekar

PG Student, Dept. of Computer Engineering, SRESCOEK, SPPU, India.

Associate Professor, Dept. of Computer Engineering, SRESCOEK, SPPU, India

ABSTRACT: Many user centered portals and websites are available now a days for sharing information and interaction some of them are Facebook, Amazon, Twitter and many more. People who are interested in any product or any service they will not only search official information but also refer user reviews on it. Due to this, online reviews, forums, portals and blogs for different product are developed and used, but how effectively the data is analyzed and exploit such huge information is a challenge. Review mining deals with extracting specific information (positive or negative) from large set of text which is written by internet users. Recent state-of-art approach such as frequency based, relation based approaches and supervised learning shows that favorable results could be obtained. It might be because patients of minority group on internet are interested in specific illnesses or drug. Instead of taking reviews from other patients, people intently importune opinions from medical professionals. However, in recent studies shows that patient's content reviews where useful in many chronic diseases and their drugs. Patients with a certain condition prefer the information shared by another patient with similar condition. The impact on patient's health was found positive by online reviews. User can see information on different drugs and also their final resultant rating based on the text reviews. User has options to browse any drug and write review on any drug they have used and based on their content, application decides the review results as positive or negative. Based on the content of the reviews, system will partition the statements and calculate the threshold of the information related to the drug.

KEYWORDS: Drug review, opinion mining, aspect mining, text mining, and topic modeling.

I. INTRODUCTION

This is a system in which, whatever information owner gets into the application are viewed by the user using the site. Every user who uses this application has to be registered and using the same registered id and password they can login into the application. User can see information on different drugs and also their final resultant rating based on the text review are seen. They can browse any drug and get information about any disease or drug. User has an option to write a review on any drug which they have used and based on their content application decides the review results whether it's a positive or negative review. Based on the content or review, system will partition the statement and work on calculating the threshold of the information related to that drug.

II. RELATED WORK

In recent years Aspect based opining mining is becoming popular. Frequency based approach extracts highfrequency noun phrases which meet the specified criteria or the constraints from the reviews as aspects. On the other hand, relation based approach identifies aspects based on the aspect – sentiment relation in the reviews. This two kind of approaches, however, may not be applicable to drug reviews as aspects are often not indicated by authers and descriptions of side effects people experiences is diverse. Moreover, grouping of the extracted phrases (noun) is another challenge, as they can't be grouped just based on semantic meanings. In contrast, topic modeling identifies aspects based on the co-occurrence of word in reviews. It has advantage that aspect identification and grouping are performed simultaneously.[2] SamanehMoghaddam and Martin Ester, they have analyzed that reviews are specific to a product or service with overall numerical rating,Opinions is a subjective belief and is the result of emotion or interpretation of facts. Then compared Aspectbased opinion mining, frequency and relation based approaches, model-based approaches and LDA basedmodels. [3] T. O'Reilly, he initiatively tried to define Web2.0 and understand its implications for the next generation of software,looking at both design pattern and business modes. Web 2.0 is the network as platform,

spanning allconnected devices; Web 2.0 applications are those that make the most of the important advantages of thatplatform, services like delivering software as a continues updates, that gets better the more people use it. [4] Bo Pang and Lillian Lee, they proposed survey which covers approaches & techniques that promise to directly enable opiningoriented information seeking system. Our focus is on methods that seek to address the new challengeswhich arise by sentiment aware application, as compared to those that are already present in moretraditional fact-based analysis. We include material on summarization of evaluative text and on boarderissues regarding privacy, manipulate and economic impact that the development of opinion-orientedinformation-access services gives rise to. To facilitate future work, a discussion of available resources,benchmark dataset and evaluation campaigns are also provided. [5] Ana-Maria Popescu and Oren Etzioni,they have illustrated on OPINE, an unsupervised information extraction system which mines reviews inorder to build a model of important product feature, their evaluation by reviewers and their relative qualityacross products. [9]Minqing Hu and Bing Liu,they aim to mine and to summarize all the customer reviews of a product. This summarization task isdifficult from traditional text summarization because they only mine the features of the product on whichthe customers have expressed their opinions, these opinions might be positive or negative. They do notsummarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews tocapture the main points as in the classic text summarization. [10] Bing Liu, Minqing Hu and Junsheng Cheng, they focused on online customer reviews of product. Two contribution are made which are, novelframework for analyzing and comparing consumer opinions of competing products. A prototype systemcalled Opinion Observer is also implemented. The system is such that with single glance of itsvisualization, the user is able to clearly see the strengths and weaknesses of each product in the mind ofcustomers in terms of various product features. The comparison is useful to both potential customers andproduct manufacturers

III. PROPOSED ALGORITHM

A. *TFIDF Steps:*

- Calculate term frequency (tf)= frequency of term t in selected document/ total no of terms in selected document
- Calculate the inverse document frequency idf - this measure the term which is less frequent
  idf(term)= log((total No. of documents)/(No. of documents with mention of term t))
- Calculate the TFIDF= tf *idf

B. *Algorithm for Drug Review Mining:*

❖ Training the data for the mining of the reviews.
1. Admin uploads a set of positive data
2. Later the keywords from these data are extracted.
3. These keywords are stored in database along with the weights & Class.
4. The class is either that the keyword is positive or negative.
5. And the weight is calculated as by the formula =>
6. Later on the weights and class and keywords are stored in database.

❖ Prediction Part
1. The reviews are selected from the database.
2. These reviews are stored in a temporary space for the prediction.
3. The keywords are detected from the review.
4. Later these keywords weight are detected from the trained values.
5. And the class count positive or negative are maintained till the end of the prediction, and finally these valuesdecide whether the post is positive or negative

IV. SIMULATION RESULTS

A. Data Set:

We use both real and synthetic data to test our proposed system. For real data dataset we took it from Online Community i.e. WebMD where lakhs of records are spread over different categories of drugs. Out of them, we retain only attributes viz. drug categories, user ids, user reviews, doctor's reviews. The attribute drug categories are taken as a sensitive attribute in the dataset. For synthetic data, according to the categories and the reviews style from the real data and the requirement of the system, synthetic data is generated.

B. Results and Analysis:

 The inputs to the system are the reviews from doctors, patients, pharmacists on a particular drug. These reviews are segregated into positive and native reviews by using the TFIDF algorithm. In here the key word which defines reviews are weighted and according to their weight the reviews are classified as positive or negative. So, it is general to say that if the number of the positive reviews is more than negative reviews, the drug is recommended and if the number of negative reviews is more than the positive reviews then another drug is recommended.

The following table 1 and figure 1 shows the recommendation of drugs and the threshold for recommendation. As the number of positive reviews is more the system will suggest the same drug, but if it's not then the system will suggest another drug from same category.

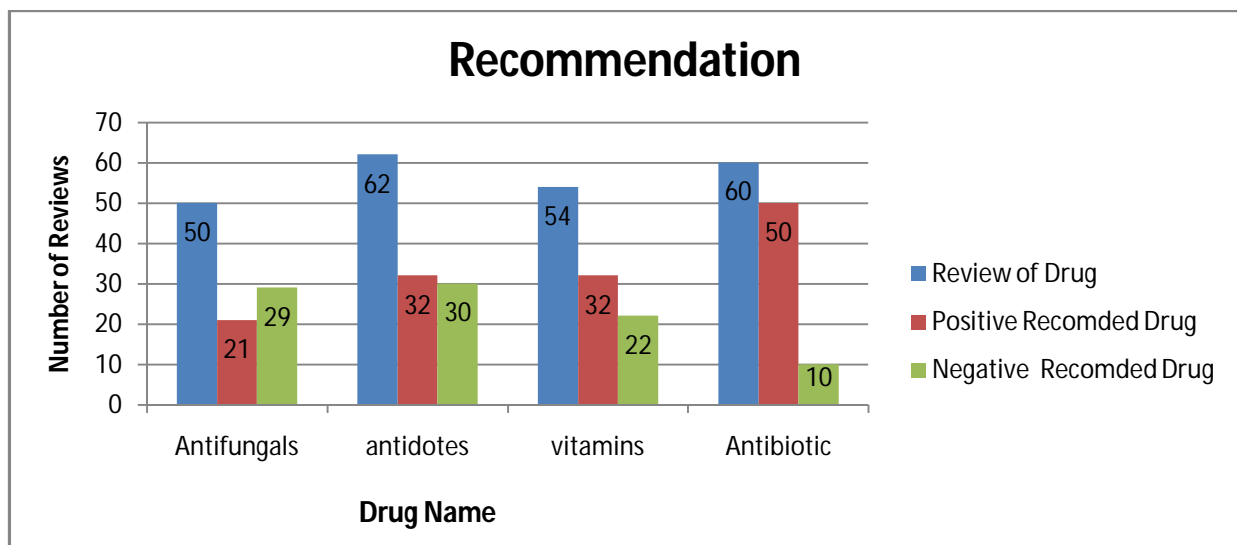| Input Drug Name	Review of Drug	Positive Reviews	Negative Reviews	Recommendation Drug
Antifungals	50	21	29	Anacine
antidotes	62	32	30	Hydroxocobalamin
vitamins	54	32	22	orange Zest
Antibiotic	60	50	10	Raw Apple Cider Vinegar

Table.1.Recommendation Table.

Fig.1.Chart Showing the Positive, Negative and Total reviews of a Drug..

After getting the required results, we still have to check the performance of the system in terms of precision, recall and F-measure. After getting these values, we can compare the current system with the previous system and can

judge the betterment of this system over the previous system. Figure 2 shows the performance evaluation in terms of Precision, Recall and F-measure.
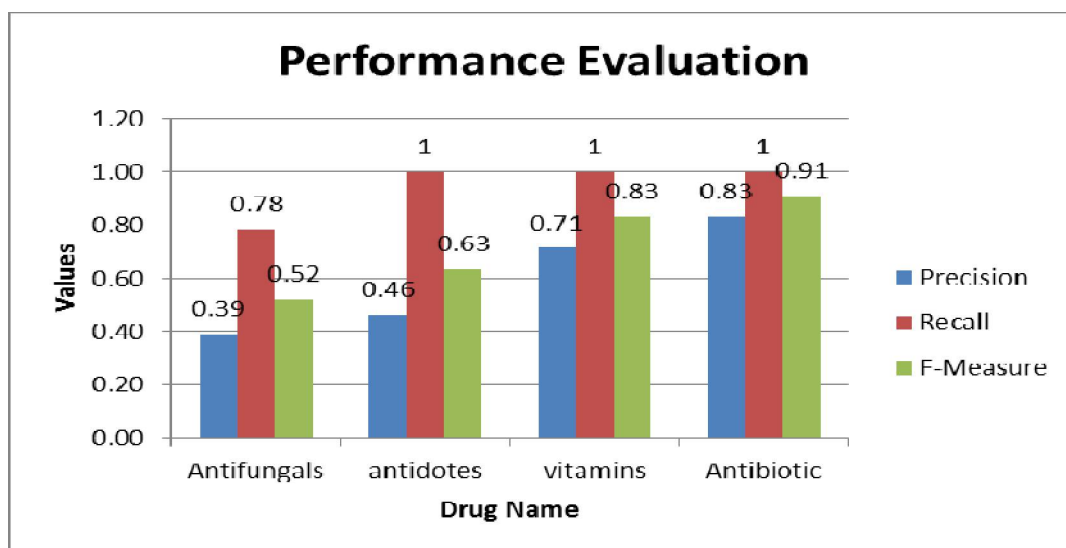

Fig. 2.Performace evaluation of the system

V. CONCLUSION AND FUTURE WORK

This system has option to initiate a message chat with doctors or pharmaceutical companies so that if they have anyqueries, they can ask them directly through this web based application. Comparing with other supervised topic modeling algorithms, this system has a unique feature that it focuses on deriving aspects from one class only, due towhich derived aspects are easier for people to interpret. It's also efficient and better as compared to previous systems. The resultant output and views will be shown on graphwhich will be generated based on the data input in the application.

REFERENCES

[1]   Victor C. Cheng, C.H.C. Leung, Jiming Liu, Fellow, IEEE, and Alfredo Milani "Probabilistic Aspect MiningModel for Drug Reviews", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26,NO. 8, AUGUST 2014.
[2]   S. Moghaddam and M. Ester, "Aspect-based opinion mining from online reviews," in Proc. Tutorial 35th Int.ACM SIGIR Conf., New York, NY, USA, 2012.
[3]   T. O'Reilly, "What is web2.0: Design patterns and business models for the next generation of software," Univ.Munich, Germany, Tech. Rep. 4578, 2007.
[4]   B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Ret., vol. 2, no. 1–2, pp. 1–135, Jan. 2008.
[5]   A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Proc. Conf. HumanLang. Technol. Emp. Meth. NLP, Stroudsburg, PA, USA, 2005, pp. 339–346.
[6]   L. Zhuang, F. Jing, and X. Zhu, "Movie review mining and summarization," in Proc. 15th ACM CIKM, NewYork, NY, USA, 2006,pp. 43–50.
[7]   Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, "Topic sentiment mixture: Modeling facets and opinions inweblogs," in Proc. 16th Int. Conf. WWW, New York, NY, USA, 2007, pp. 171–180.
[8]   D. Giustini, "How web 2.0 is changing medicine," BMJ, vol. 333, no. 7582, pp.1283–1284, 2006.
[9]   M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. KDD,Washington, DC, USA, 2004, pp. 168–177.
[10]   B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comaring opinions on the web," in Proc. 14[th]Int. Conf. WWW, New York, NY, USA, 2005, pp. 342–351.
[11] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in Proc. 18[th]ACM CIKM, New York, NY, USA, 2009, pp. 375–384.

BIOGRAPHY

Abhimanyu Dnyandeo Sangale is a PG Student **Prof. P. N. Kalavadekar** is Associate Professor and PG Co-ordinator in the Computer Engineering Department, College of Engineering, Kopargaon SPPU, Pune. He received Bachelor of Engineering (BE) degree and pursuing Master's Degree in SRES COEK, SPPU, Pune India. His research interests are Data Mining, Algorithms, IOT etc.