

A SURVEY ON DRUG REVIEWS USING DATA MINING MODEL

Mr Sangale Abhimanyu D.¹, Prof.P.N.Kalavadekar²

¹ PG Student, SRES'COE, Kopargaon, SPPU, Maharashtra, India

² Associate Professor, SRES'COE, Kopargaon, SPPU, Maharashtra, India

ABSTRACT

Many user-centred portals and websites are available now a day for sharing information and interaction; some of them are Amazon, Facebook, 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 analysed 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 favourable 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 patients 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 patients 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.

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

1. 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.

2. LITERATURE SURVEY

In recent years Aspect based opinion mining is becoming popular. Frequency based approach extracts high frequency 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 authors 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.

2.1 Samaneh Moghaddam and Martin Ester [2]

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 Aspect based opinion mining, frequency and relation based approaches, model-based approaches and LDA based models.

2.2 T. O'Reilly [3]

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 all connected devices; Web 2.0 applications are those that make the most of the important advantages of that platform, services like delivering software as a continues updates, that gets better the more people use it.

2.3 Bo Pang and Lillian Lee [4]

They proposed survey which covers approaches & techniques that promise to directly enable opining oriented information seeking system. Our focus is on methods that seek to address the new challenges which arises by sentiment aware application, as compared to those that are already present in more traditional fact-based analysis. We include material on summarization of evaluative text and on boarder issues regarding privacy, manipulate and economic impact that the development of opinion-oriented information-access services gives rise to. To facilitate future work, a discussion of available resources, benchmark dataset and evaluation campaigns is also provided.

2.4 Ana-Maria Popescu and Oren Etzioni [5]

They have illustrated on OPINE, an unsupervised information extraction system which mines reviews in order to build a model of important product feature, their evaluation by reviewers and their relative quality across products.

2.5 Minqing Hu and Bing Liu [9]

They aim to mine and to summarize all the customer reviews of a product. This summarization task is difficult from traditional text summarization because they only mine the features of the product on which the customers have expressed their opinions, these opinions might be positive or negative. They do not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization.

2.6 Bing Liu, Minqing Hu and Junsheng Cheng [10]

They focused on online customer reviews of product. Two contribution are made which are, novel framework for analyzing and comparing consumer opinions of competing products. A prototype system called Opinion Observer is also implemented. The system is such that with single glance of its visualization, the user is able to clearly see the strengths and weaknesses of each product in the mind of customers in terms of various product features. The comparison is useful to both potential customers and product manufacturers.

3. SYSTEM OVERVIEW

In this system for Drug Reviews using Data Mining Model, more specially following steps are performed:

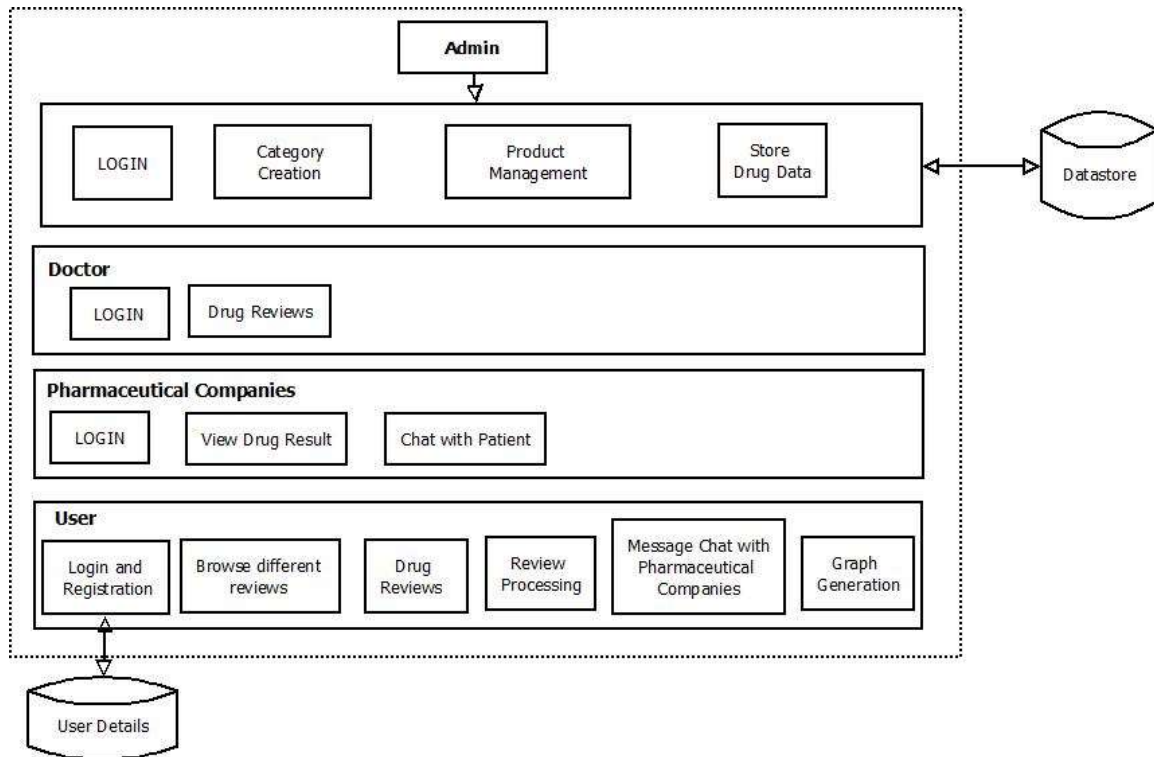


Fig -1: System Architecture

3.1 Algorithm which is implement in drug mining.

Tfidf Steps

Calculation

1. Calculate term frequency (tf)= frequency of term t in selected document/ total no of terms in selected document
2. Calculate the inverse document frequency idf - this measure the term which is less frequent

$$\text{idf}(\text{term}) = \log\left(\frac{\text{total No. of documents}}{\text{No. of documents with mention of term t}}\right)$$
3. Calculate the TFIDF= tf *idf

Algorithm for Drug Review Mining:

- 1) Training the data for the mining of the reviews.
 - Ø Admin uploads a set of positive data
 - Ø Later the keywords from these data are extracted.
 - Ø These keywords are stored in database along with the weights & Class.
 - Ø The class is either that the keyword is positive or negative.
 - Ø And the weight is calculated as by the formula =>
 - Ø Later on the weights and class and keywords are stored in database.
- 2) Prediction Part
 - Ø The reviews are selected from the database.
 - Ø These reviews are stored in a temporary space for the prediction.
 - Ø The keywords are detected from the review.
 - Ø Later these keywords weight are detected from the trained values.
 - Ø And the class count positive or negative are maintained till the end of the prediction, and finally these values decide whether the post is positive or negative.

The proposed system contains different modules which work on reviews as follows:

- **Admin Login:** Admin or the site owner will login into the application with help of their credentials and they have complete right to manage and populate the information in the site through the panel.
- **Category creation and product management:** Admin has the right to create the category of diseases or domain and under that they can store or feed the drug data and the information. Whatever information owner gets into the application is viewed by the user using the site.
- **User registration and login:** Every user who uses this application has to be registered and using the same registered id and password they can login into the application.
- **Browse different reviews:** User once login 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.
- **Reviews from user:** 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 its a positive or negative review.
- **Review processing:** Based on the content or review, system will partition the statement and work on calculating the threshold of the information related to that drug.
- **Doctor login:** Doctor has a login in which they can also give their own review on any particular drug.
- **Pharmaceuticals login:** Pharmaceuticals companies get a complete view of the drug result which they can use for further processing and refining their own product. They can also chat with patient to help them out.
- **Message Module:** User has an option to initiate a message chat with pharmaceutical companies so that if they have any queries they can ask them directly through this web based application.
- **Graph Generation:** The resultant output and views will be shown on graph which will be generated based on the data input in the application.

4. CONCLUSIONS

This system has option to initiate a message chat with doctors or pharmaceutical companies so that if they have any queries they can ask them directly through this web based application. Comparing with other supervised topic modeling algorithms, PAMM has a unique feature that it focuses on deriving aspects from one class only, due to which derived aspects are easier for people to interpret. The resultant output and views will be shown on graph which will be generated based on the data input in the application.

5. ACKNOWLEDGEMENT

I want to thanks to my guide Prof.P.N.Kalavadekar for his important guidance and encouragement in every situation. Suggestions given by him broaden my vision and guided me to succeed in this work. I am also very grateful for his guidance and comments while designing part of my research paper and learnt many things under his leadership.

6. REFERENCES

- [1] Victor C. Cheng, C.H.C. Leung, Jiming Liu, *Fellow, IEEE, and Alfredo Milani* "Probabilistic Aspect Mining Model 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. Human Lang. 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, New York, 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 in weblogs," 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 comparing opinions on the web," in Proc. 14th 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. 18th ACM CIKM, New York, NY, USA, 2009, pp. 375–384.

