Proceedings of
INTERNATIONAL CONFERENCE ON INNOVATIVE TRENDS IN ENGINEERING RESEARCH (ICITER-2016)
International Journal of Innovations in Engineering, Research and Technology,IJIERT-ICITER-16,ISSN:2394-3696
26th June,2016

A REVIEW ON: SENTIMENT ANALYSIS FOR DRUG USING USER REVIEWS

[PAPER ID ICITER-C115]

SUPRIYA PHATAK
Annasaheb Dange College of Engineering, Sangli, India
Email: supriya4268@gmail.com

S.V. PATIL
Annasaheb Dange College of Engineering, Sangli, India

RUPALI PATIL
Annasaheb Dange College of Engineering, Sangli, India

NITINKUMAR ANEKAR
MIT College of Engineering, Pune, India
Email:nitin.anekar@gmail.com

ABSTRACT:

Merchants selling products on the Web often ask their customers to review the products that they have purchased and the associated services. As e-commerce is becoming more and more popular, the number of customer reviews that a product receives grows rapidly. For a popular product, the number of reviews can be in hundreds or even thousands. This makes it difficult for a potential customer to read them to make an informed decision on whether to purchase the product. It also makes it difficult for the manufacturer of the product to keep track and to manage customer opinions. For the manufacturer, there are additional difficulties because many merchant sites may sell the same product and the manufacturer normally produces many kinds of products. In this research, we aim to mine and to summarize all the customer reviews of a product. This summarization task is different from traditional text summarization because we only mine the features of the product on which the customers have expressed their opinions and whether the opinions are positive or negative. We 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. Our task is performed in three steps: (1) mining product features that have been commented on by customers; (2) identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative; (3) summarizing the results. This paper proposes several novel techniques to perform these tasks. Our experimental results using reviews of a number of products sold online demonstrate the effectiveness of the techniques.

KEYWORDS: Aspect Detection, Sentiment Analysis, Topic Modeling

INTRODUCTION:

Electronic media are increasingly used to obtain medical information and advice. Health information on the Internet ranges from personal experiences of medical conditions and patient discussion groups to peer reviewed journal articles and clinical decision support tools. A study on how consumers in America search for health-related information1 shows that the Web is the most widely used resource for health information. Nevertheless, finding the best knowledge source to comply a specific information need is difficult, because relevant information can be either hidden in web pages or encapsulated in social media data such as blogs and Q&A portals. Through content analysis, this paper tries to give an overview on content differences in the various social media resources on health-related topics. We focus on health-related information provided in the Internet for two reasons. First, health-related experiences and medical histories offer unique data for research purposes, for practitioners, and for patients. Second, it is still an open question whether existing text and content analysis tools are able to process medical social media data and to identify relevant (medical) information out of them.

Online reviews, blogs, and discussion forums such as WebMD on chronic diseases and medicines are becoming important supporting resources for patients. Extracting useful information from these substantial bodies is very difficult and challenging. Opinion mining

or sentiment analysis involves the extraction of useful information (e.g., positive or negative sentiments of a product) from a large quantity of text opinions or reviews authored by Internet users interest, and thus it will help us to improve the web site structure. We decompose the problem of review mining into the following main subtasks:

1. Identify product features.
2. Identify opinions regarding product features.
3. Determine the polarity of opinions.
4. Rank opinions based on their strength.

With the advent of Web people are enabled and encouraged to contribute their contents to the Internet. Many user-centered platforms are now available for information sharing and user interaction, such as Epinion, Amazon, Facebook and Twitter. Nowadays when people are interested in a product or a service, they usually not only look for official information from product manufacturers or service providers, experienced and practical opinions from the customers' and users' points of view are also influential. As a result, online reviews, blogs and forums dedicated for different kinds of products are pervasive, and how to effectively analyze and exploit such immense online information source is a challenge. While medical professional's viewpoint is the dominating factor in the provision of healthcare services, there are constant concerns over the patients' role in their treatment. According to a recent study of the discrepancy between clinician's interpretation of patients' symptoms and patients reported drug usage experience, it is found that patients' reports are more highly concordant with overall health status than clinician's report. The study points out that self report are more sensitive to underlying changes of the patient's functional status. The rapid development of social networking applications generate tremendous amount of information which can be extracted for enhancing the understanding of products and their relationships with people. Opinion mining for healthcare has focused on polarity classification, and a wider spectrum of affective states in healthcare has been analyzed.

Opinion mining (or sentiment analysis) deals with the extraction of specified information (e.g., positive or negative sentiments of a product) from a large amount of text opinions or reviews authored by Internet users. In many situations, solely an overall rating for a review cannot reflect the conditions of different features of a product or a service. For instance, a camera may come with excellent image quality but poor battery life. As a result, more sophisticated aspect level opinion mining approaches have been proposed to extract and group aspects of a product or service and predict their sentiments. Previous studies of opinion mining usually deal with popular consumer products or services such as digital cameras, books, electronic gadgets, etc. Nevertheless, recent studies have shown that patient generated contents are useful and important, especially for chronic diseases and drugs with afflicting side effects. Many patients hope to get more information from other patients with similar conditions. More importantly, authors sometimes do not indicate which aspects they are describing, they just give descriptions of symptoms, feelings and comments. The following summarizes the features of drug reviews.

- Drug reviews have a small number of kinds of aspects: price, ease of use, dosage, effectiveness, side effects and people's experiences. Aspects are usually not mentioned explicitly.
- Descriptions of effectiveness, side effects and people's experiences are diverse.
- Side effect and effectiveness descriptions are different from drug to drug.

LITERATURE SURVEY:

Bo Pang et al. [1] provides an important part of our information-gathering behavior has always been to find out what other people think, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as direct response to the surge of interest in new systems that deal directly with opinions as a first-class object. *Victor C. Cheng* [5] drug reviews from patient are documented on on-line but mining significant topics is very challenging. Interpretation of patient symptoms and drugs usage are used to make clinical report the study of this point is more sensitive to view functional status of patient. Opinion mining focuses on polarity classification another approach of review is based on computation of mutual information. Non negative matrix factorization recent advancement of NMF is similar to that of K-means algorithm. In this paper Regression Probabilistic Principal Component Analysis (RPPCA) was introduced to review sentiment values and also explore how to medical data has been used for document analysis. *Simon Lacoste- Julien et. Al.*[6] given probabilistic method as became very important for dimensionality reduction for text or image

documents. Dimensionality reduction learning is often necessary because of data analysis. Principal Components Analysis (PCA) and Fisher Discriminate Analysis (FDA) is important learning algorithm for discriminative learning. This paper discusses on alternative method for finding reduced dimensionality representation on a discriminative frame work. DisLDA , a Discriminative Variation on Latent Dirichlet Allocation (LDA) a dependent linear transformation for dimensionality reduction and classification. *Wei Jin et al* [7] given combination of text data and document metadata are viewed because of Bayesian multinomial mixture models like Latent Dirichlet Allocation (LDA) which makes text analysis simple, use of reduces the dimensionality of data and able to describe interpretable and semantically coherent topics are basically text data was accompanied by metadata such as dates, about authors and publication. Currently for specifying to generative model and implementing model has been developed.

IMPLEMENTATION:

Opinion mining (or sentiment analysis) deals with the extraction of specified information (e.g., positive or negative sentiments of a product) from a large amount of text opinions or reviews authored by Internet users. In many situations, solely an overall rating for a review cannot reflect the conditions of different features of a product or a service.

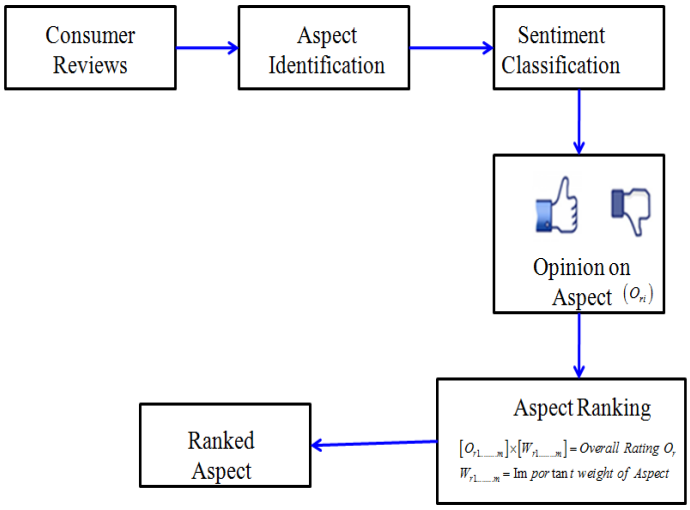

Fig. No. 1: System Architecture

For instance, a camera may come with excellent image quality but poor battery life. As a result, more sophisticated aspect level opinion mining approaches have been proposed to extract and group aspects of a product or service and predict their sentiments or ratings. This paper gives address the opinion mining problems for drug reviews. As many drug review websites are equipped with rating functions, prediction of sentiments is not the task. Instead a model for identifying a set of aspects relating to class labels or meta-information of drug reviews is proposed. For example, if the reviews are associated with age information, people may be interested in studying the aspect difference between ages of patients.

ALGORITHMIC STEPS:

PAMM (Probabilistic Aspect Mining Model)

Drug reviews have a small number of kinds of Aspects: price, ease of use, dosage, effectiveness, side effects and people's experiences, age.

Aspects are usually not mentioned explicitly

Descriptions of effectiveness, side effects and people's experiences are diverse.

Side effect and effectiveness descriptions are different from age to age.

Requires:

$\{(x_n)\}_n^N = 1$, *where* $x_n \in R^M$, $y_n \in \{0,1\}$

K: Number of Aspects to be derived,

σ^2 : Variance of noise

c : Parameter of the logistic function,

δ : Threshold for stopping EM iteration.

The program flow is given as:

1. Compute the empirical mean for $\{(x_n)\}_n^N = 1$ $(i.e.\ \mu)$
2. Center the data by $x_n \leftarrow (x_n - \mu)$ for *n=1,…..,N,*
3. Initialize the entries of W randomly to small positive numbers.
4. **Repeat**
5. {E-step}
6. **for** *n=1 to N* **do**
   Calculate
   end for
7. {M-step}
8. **for** *i=1 to M* **do**
   Update W_i
   end for
9. **until** Change of $\|W\|_{Frob}$ in consecutive EM iterations < δ
10. **return** *W*

CONCLUSION AND FUTURE SCOPE:

Opinion mining is a field of study which helps to extract aspects from the opinions of the internet users. As the internet data contains large amount of data, designing and implementing opinion mining algorithms is difficult and very complex. Comparing with different supervised topic modeling algorithms, PAMM contains a distinctive feature that it focuses on explanation aspects for one category solely. This feature reduces the opportunities of forming aspects from reviews of various categories and thus the derived aspects area unit easier for folks to interpret.

REFERENCES:

i. B. Pang and Lillian Lee, "*Opinion mining and sentiment analysis*", Found. Trends Inf. Ret., Vol. 2, No. 1–2, Jan. 2008.pp. 1–135

ii. Gayathri R. Krishna, Kavitha , Yamini ,Rekha, "*Analysis of Various Opinion Mining Algorithms*," International Journal of Computer Trends and Technology, Vol. 22, No.2,April 2015.

iii. V. Ranjani Gandhi, N. Priya, "*Literature Survey on Data Mining and Statistical Report for Drugs Reviews*," International Journal of Innovative Research in Computer and Communication Engineering, Vol.3, Issue 3, March 2015.

iv. Satchinathan P., Dr. T. Stephen Thangaraj, R. Nandakumar, "*Automatic Integration of Drug Indications from Multiple Health Resources*," International Journal of Innovative Research in Computer and Communication Engineering, Vol.3, Issue 3, March 2015.

v. Victor C. Cheng, Leung, Jiming Fellow, "*Drug Review Mining with Regression Probabilistic Principal Component Analysis*", HI- KDD'12, Beijing, China, 2012

vi. Simon Lacoste- Julien, Michael I. Jordan, "*Dis LDA: Discriminative Learning for Dimensionality Reduction and Classification*," L3S Research Center, University of Hannover Germany,2010.

vii. Wei Jin, Hung Hay Ho, and Rohini K.Srihari, "*Opinion Miner: A Novel Machine Learning System for Web Opinion and Extraction*", KDD'09, 2009 .

viii. David Mimno, Andrew McCallm, "*Topic Models Conditioned Arbitrary Features with Dirichlet – Multinomial Regression*", UAI-P-2008-PG-411-418,2012.