Opinion Extraction of Drug Reviews Using PAMM Model

Adhav Pavan D.[1], Pardeshi Gaurav S.[2], Bhad Sainath A.[3], Patel Nishit B.[4]
[1] B.E Comp Engineering, S.N.D COE & RC, YEOLA, Maharashtra, India
[2] B.E Comp Engineering, S.N.D COE & RC, YEOLA, Maharashtra, India
[3] B.E Comp Engineering, S.N.D COE & RC, YEOLA, Maharashtra, India
[4] B.E Comp Engineering, S.N.D COE & RC, YEOLA, Maharashtra, India

ABSTRACT

By the drastic popularity of Internet, the study of online reviews through blogs, discussion forums etc., have become most popular way for the patients to have medicines for chronic diseases. This paper consists of various research parameters for opinion extraction of drug reviews and techniques used in it. A probabilistic theory is developed to extract useful information from those reviews, called as PAMM (Probabilistic Aspect Mining Model). PAMM has a particular characteristic that it concentrates on finding opinion aspect related to one class instead of finding aspect for all classes simultaneously on each execution. This reduces the chances of mixed concepts of other classes. The aspects found are also responsible to differentiate a class from other classes. The paper gives idea to propose an efficient EM algorithm to propose opinion aspects for various groups of ages. An EM algorithm is used for finding approximate parameters of an underlying distribution from data set when it has missing values.

Keyword: - *Drug Review, Opinion Extraction, Aspect Mining, Probabilistic Aspect Mining Model (PAMM), Text Mining.*

---

1. INTRODUCTION

Nowadays many user-centered platforms are available for information sharing and user interaction, such as Blogs, Facebook, and Twitter etc., which act as an interface and to which people are connected and share their opinions over the internet. 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; opinions from the active customers' and users' points of view are also influential. Opinion mining (or sentiment analysis) [1], [6], [14]-[16] 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. 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 [1], [14], [17]-[21].

Previous studies of opinion mining usually deal with popular consumer products or services such as digital cameras, books, electronic gadgets, etc. Entities of medical domain are of far less concerned. It may be because patients are minority groups on the Internet and they are only concerned with specific illnesses or drugs that they are experiencing. Furthermore, people tend to solicit opinions from medical professionals rather than patients. Nevertheless, recent studies have shown that patient generated contents are useful and important [1], [22], [8]-[10], especially for chronic diseases and drugs with afflicting side effects. Many patients hope to get more information from other patients with similar conditions. They can also share their experience and propose practical ways to alleviate symptoms and side effects of drugs. These online communities were found to have positive impacts on patient health [1], [11]-[13].

Unlike general products or services, drugs have a very limited number of kinds of aspects: price, ease of use, dosages, effectiveness, side effects and people's experiences. There are other more technical aspects such as chemical or molecular aspects, but they are almost not mentioned in drug reviews. A difficulty in dealing with drug reviews is that the wording in describing effectiveness, side effects and people's experiences are very diverse. In particular, side effects are drug dependent: a set of side effect symptoms for a drug is very unlikely applicable to another drug. This impedes some opinion mining approaches based on lexicons. More importantly,

authors sometimes do not indicate which aspects they are describing, they just give descriptions of symptoms, feelings and comments [1].

In this paper we address opinion mining problem for drugs and proposed a novel Probabilistic Aspect Mining Model (PAMM) for mining the drug reviews with structured information and many of the drug review websites are managed to perform sentiment opinion mining and grading functions but they tend to produce labeled information as the extracted topic is useful for patients because they can study about various aspects of the drugs and their functionalities [1].

The task is different from general aspect-based opinion mining in which the task aims to extract all aspects and their sentiments from reviews. Referring to the problem definition, not all the aspects but only relevant aspects need to be extracted. Sometimes, an aspect may need to be segmented further (in finer granularity) because only limited components of it are required. For instance, considering the aspect of side effects of a drug, male patients may be anxious about a specific side effect while other side effects are of less concerned [1].

We propose a novel probabilistic aspect mining model (PAMM) to mine the aspects of drug reviews correlated with categorical information. This can be regarded as a topic model with the derived topics treated as aspects. The proposed model is very useful to patients and pharmaceutical companies because various aspects of a drug can be identified. In addition, the results can be used to compile sentiment lexicons for drug reviews. Words of aspects correlating with high satisfactory ratings can be regarded as positive sentiment words and vice versa. Practically, this model is not limited to drug reviews. It can be applied to other domains such as product reviews and service reviews for studying aspects pertaining to different groupings of reviews [1].

2. LITERATURE SURVEY

Given a corpus of reviews (assume every review is in bag-of-words format and a class label is assigned), words highly correlated with the class label can be identified by many approaches such as class conditional probability of words, information gain [1], [23], association rules [1], [24], pointwise mutual information (PMI) [1], [25], etc. These approaches, unfortunately, suffer from a severe problem: it is difficult to understand the underlying aspects or concepts from just a set of words correlated with a class label. There is no intuitive algorithms to group the words so that each group conveys one or a few easily understandable concepts.

Aspect-based opinion mining is becoming popular in recent years. Frequency based approach [1], [15] extracts high frequency noun phrases which meet the specified criteria or constraints from the reviews as aspects. On the other hand, relation based approach [1], [19], [26] identifies aspects based on the aspect-sentiment relation in the reviews. These two kinds of approaches, however, may not be applicable to drug reviews as aspects are often not indicated explicitly by authors and descriptions of side effects and people's experiences is diverse. Moreover, grouping of the extracted noun phrases is another challenge as they cannot be grouped just based on semantic meanings. In contrast, topic modelling identifies aspects based on the co-occurrence of words in reviews. It has an advantage that aspect identification and grouping are performed simultaneously[1].

Topic modeling [1], [27]–[32] (e.g., LDA [27]) is a popular probabilistic approach in understanding a corpus. With this approach, a set of topics, which are represented by multinomial distributions over vocabulary words, are inferred. When the words of a topic is sorted according to the probabilities, high probability words of a topic are usually semantically correlated and the concept or aspect of the topic can be captured manually. For example, Topic Sentiment Mixture (TSM) [1], [17], Joint Sentiment/Topic (JST) [1], [20] model and Aspect and Sentiment Unification Model (ASUM) [1], [33] were proposed to extract both the aspects and predict their associated sentiments. Nevertheless, these aspectbased opinion mining methods may not be appropriate to address the problem defined in the previous section as the extracted aspects may not be related to the specified class labels and the performance depends on the manual selection of seed words [1].

3. PROPOSED SYSTEM

PAMM has a unique feature that is it focuses on finding aspects relating to one class only rather than finding aspects for all classes simultaneously in each execution. This reduces the chance of having aspects formed from mixing concepts of different classes; hence the identified aspects are easier to be interpreted by people. The aspects found also have the property that they are class distinguishing they can be used to distinguish a class from other classes. Aspect-based opinion mining is becoming popular in recent years. Frequency based approach extracts high frequency constraints from the reviews aspects. On the other hand, relation based approach identifies aspects. These two kinds of approaches, however, may not be applicable to

drug reviews as aspects are often not indicated explicitly by authors and descriptions of side effects and people's experiences are diverse. So, the opinion mining problems for drug reviews are addressed. As many drug review websites are equipped with rating functions. Instead a model for identifying a set of aspects relating to class labels or meta-information of drug reviews is proposed by using PAMM [5].

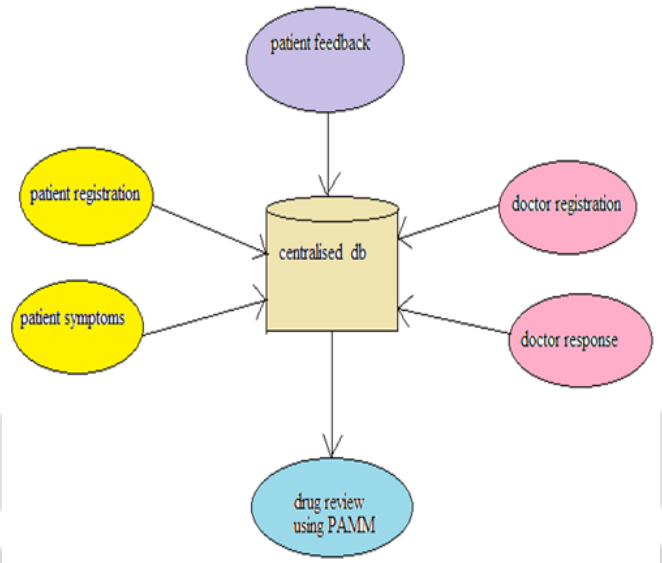

Fig1. System Flow for drug review.

This model is used to find aspects relating to different segmentation of data such as different age groups, gender, drug name and drug formula. Patients register and login. Doctor registers and login. Patients and Doctors details are stored in centralised database. Now patient logins and enter the symptoms. Symptoms are stored in database. Doctor will login and give response to the patients. Patient will check response and gives feedback. Patient can search drug review based on age, gender, drug name and drug formula [5].

A. Algorithm for PAMM:

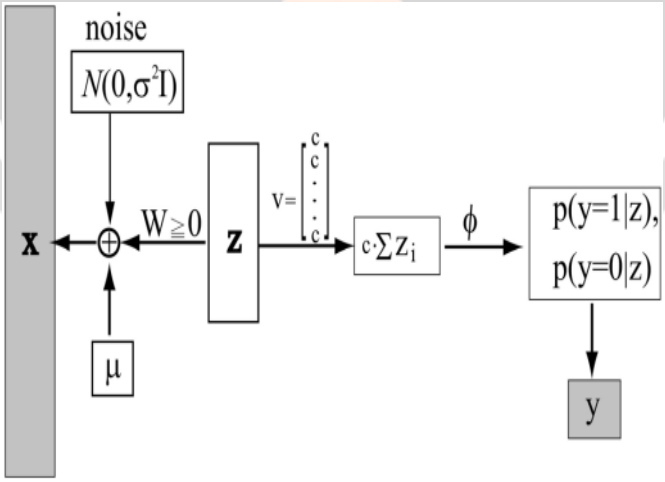

Fig2. PAMM for generating observed data x and label y from latent variable z.

Referring to the figure, data points and the associated class labels are generated as follows.

- Draw**z** ~ N(0, **I**)
- Draw**x** ~ N(**Wz** + μ, **I**);
- Draw y ~ (p(y = 0|**z**), p(y = 1|**z**)),

Where μ is the mean of the observed data, is the Gaussian noise level on x, W ∈ is a matrix having non-negative entries, $p(y = 1|\mathbf{z})$ and $p(y = 0|\mathbf{z})$ are Probabilities.

B. PAMM Algorithm:

1: Compute the empirical mean for $\{(x_n)\}^N_n = 1$ (i.e. μ).
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
7: Calculate z^*_n
8: end for
9: {M-step}
10: for i=1 to M do
11: Update W_i ,
12: end for
13: until Change of ||W|| in consecutive EM iterations < δ
14: return W.

C. Architecture

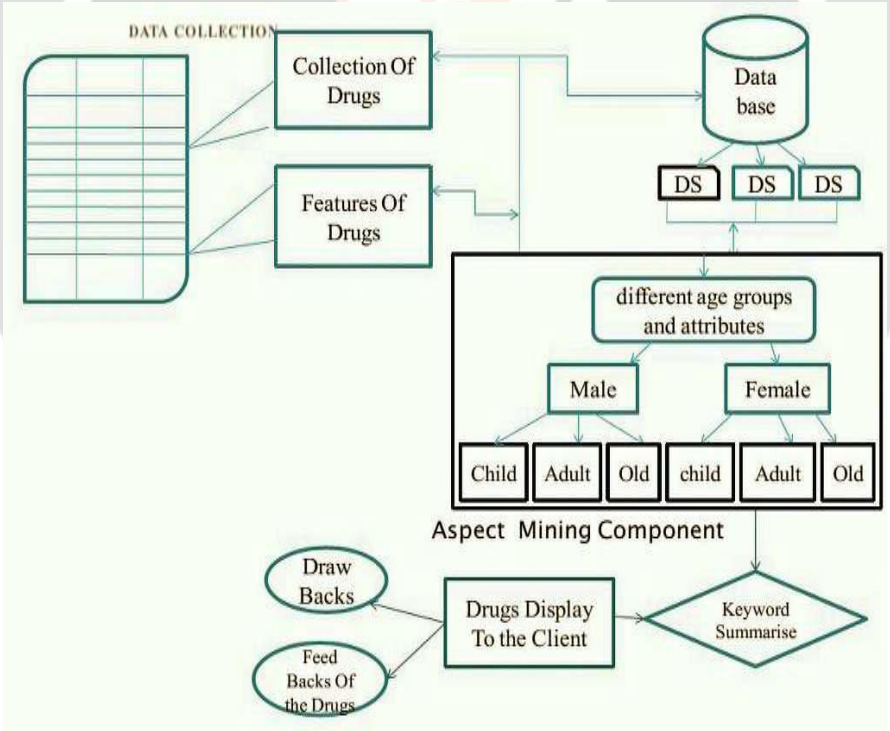

Fig.3. Aspect mining architecture

A. COLLECTION OF DRUGS

The collection of Drugs called as Drug database for collecting drugs that are highly accessible with broadly covering quality and contains the information about drugs and drug targets in online community as bioinformatics and chemical informatics resources are gathered together for the detailed drug reviews that is Pharmaceutical and chemical and pharmacological and information about comprehensive data also been collected.

As there is a high scope for comprehensive referencing and data detailed description it is more akin to the drug database that is widely used by the medicinal chemists or physicians or drug industry or student and general public as the collection of drug database are used to discover and repurpose drugs which exist and newly discovered illnesses as the latest drug database contains entries which are approved by the small molecule drug factors [2].

B. FEATURES OF DRUGS:

All the existing drug features can be classified into two different ways like Physiological Tolerance and Drug tolerance which is subjected to the reaction when there is a particular drug progressively decreasing day by day then their needs concentration growth for enabling desired effects.

Due to the need of drug development mainly depends on the rate on the particular drug or frequency of use of drugs and differential development of the same drugs for improving drugs we have to increase required parameters with the same magnitude of responsibility.

Due to change in neurology in the frequent drug review leads to the changes in receptor desensitization and depletion in the neural [l] transmitters this process helps in neural adaptation development environment [2].

C. DATABASE

Every organization collects the data is called database abd typical organizational model for the database aspects reality makes a supporting process for requiring information like availability of the required data events.

As the database management system is the computer software applications which are used to interact with the different type of user and other application formats for analyzing the data as a DBMS is designed in the following format for defining the data for creating new data type and to implement query processing then updating and administration process [2].

D. WORKING PRINCIPLES

The set of Drugs are extracted from the common database storage where all the data being stored in the particular data storage server and the feature of the drug for the patient benefit is saved in the same storage. As the data extracted from the common database is collected and stored on another medical database server for the classification for a user to view the information about any particular drug for the drug clarifications.

In the Aspect Mining Component we described different age, gender group and attribute components for the experimental work formats for extracting the information from the medical database, we classify two categories such as male and female types with three different kinds of attributes like child then adult and old [2].

4. CONCLUSION

Online reviews are playing significant role with consumers when they look for purchasing products. This provides all the positive, negative and neutral effects customer reviews can have on business. Customer reviews and statements are becoming powerful in a time when the consumer is trusting fellow consumers. PAMM is used to find adverse effect of drugs. It has a unique feature that it focuses on deriving aspects for one class only. PAMM uses all the reviews and finds the aspects that are helpful in identifying the target class. As human lifespan becomes longer and living environment becomes increasingly polluted, medical domain data mining becomes one of the focused research areas. PAMM is proposed for mining aspects relating to specified labels or groupings of drug reviews. Drug reviews can be searched based on age, gender, name and formula.

REFERENCES

[1] Victor C. Cheng, C.H.C. Leung, Jiming Liu, Fellow, IEEE, and Alfredo Milani ," Probabilistic Aspect Mining Model for Drug Reviews", Hong Kong Baptist University, Kowloon 1234, Hong Kong.

[2] V.Ranjani Gandhi, N.Priya "A Novel Based Statistical Information for Medicine Reviews Using Data Mining Concepts" International Journal of Innovative Research in Science, Engineering and Technology Vol. 4, Issue 3, March 2015.

[3] Kiran G1, J. Laxminarayana "Feasible Aspect Mining Model for Drug Reviews" Aurora's International Journal of Computing 2015 Vol. 2. Issue 2.

[4] 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.

[5] Amtul Kubra, Ayesha Khan, Saleha Farha "Modelling A Probabilistic Theory for Mining The Reviews on Pharmaceuticals" International Journal of Advanced Technology and Innovative Research Vol.07,Issue.06, July-2015.

[6] B. Pang and L .Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Ret.*, vol. 2, no.1–2, pp. 1–135, Jan. 2008

[7] 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.

[8] K. Denecke and W. Nejdl, "How valuable is medical social media data? content analysis of the medical web," *J. Inform. Sci.*, vol. 179, no. 12, pp. 1870–1880, 2009.

[9] X. Ma, G. Chen, and J. Xiao, "Analysis on an online health social network," in *Proc. 1st ACM Int. Health Inform. Symp.*, New York, NY, USA, 2010, pp. 297–306.

[10] A. Névéol and Z. Lu, "Automatic integration of drug indications from multiple health resources," in *Proc. 1st ACM Int. Health Inform. Symp.*, New York, NY, USA, 2010, pp. 666–673.

[11] J. Leimeister, K. Schweizer, S. Leimeister, and H. Krcmar, "Do virtual communities matter for the social support of patients? Antecedents and effects of virtual relationships in online communities," *Inform. Technol. People*, vol. 21, no. 4, pp. 350–374, 2008.

[12] R. Schraefel, R. White, P. André, and D. Tan, "Investigating web search strategies and forum use to support diet and weight loss," in *Proc. 27th CHI EA*, New York, NY, USA, 2009, pp. 3829–3834.

[13] J. Zrebiec and A. Jacobson, "What attracts patients with diabetes to an internet support group? A 21-month longitudinal website survey," *Diabetic Med.*, vol. 18, no. 2, pp. 154–158, 2008.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comaring opinions on the web," in *Proc. 14th Int. Conf. WWW*, New York, NY, USA, 2005, pp. 342–351.

[20] 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.

[21] I. Titov and R. McDonald, "A joint model of text and aspect ratings for sentiment summarization," in *Proc. 46th Annu. Meeting ACL*, 2008, pp. 308–316.

[22] J. Sarasohn-Kahn, "The wisdom of patients: Health care meets online social media," California Healthcare Foundation, Tech. Rep., 2009.

[23] T. Mitchell, *Machine Learning*. Boston, MA, USA: McGraw-Hill, 1997.

[24] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. VLDB*, San Francisco, CA, USA, 1994, pp. 487–499.

[25] C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999.

[26] S. Baccianella, A. Esuli, and F. Sebastiani, "Multi-facet rating of product reviews," in *Proc. 31st ECIR*, Berlin,, Germany, 2009, pp. 461–472.

[27] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Jan. 2003

[28] .T. Griffiths, M. Steyvers, D. Blei, and J. Tenenbaum, "Integrating topics and syntax," in *Proc. Adv. NIPS*, 2005, pp. 537–544.

[29] D. Blei and J. Lafferty, "Correlated topic models," in *Proc. Adv. NIPS*, 2006, pp. 147–154.

[30] A. McCallum and X. Wang, "Topic and role discovery in social networks with experiments on enron and academic email," *J. Artif. Intell. Res.*, vol. 30 no. 1, pp. 249–272, 2007.

[31] D. Putthividhya, H. Attias, and S. Nagarajan, "Independent factor topic models," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, New York, NY, USA, 2009, pp. 833–840.

[32] M. Rosen-Zvi, C. Chemudugunta, T. Griffiths, P. Smyth, and M. Steyvers, "Learning author-topic models from text corpora," *ACM TOIS*, vol. 28, no. 1, pp. 1–38, Jan. 2010.

[33] Y. Jo and A. Oh, "Aspect and sentiment unification model for online review analysis," in *Proc. 4th ACM Int. Conf. WSDM*, New York, NY, USA, 2011, pp. 815–824.