

Smart Recommendation System for Off-the Shelf Medicines

Ratna Sathappan^{1*}, Tholu Sai Indira², A. Meenapriyadarsini³

^{1,2}Undergraduate student, ³Assistant Professor, CSE department, SRM University, Kattankulathur
*Corresponding Author Email: ratnasathappan@gmail.com

Abstract

Internet usage has been at an all-time high from 2000's vintage years. The people who have access to the internet use it for numerous reasons such as social networking, marketing, promoting, enhancing businesses, consultancy, research, gaming and the list goes on. In the recent years, Review websites have flourished, where people share their opinion about a product, with an increase in response rate and reliability. Recommendations are made by mining data from review websites. Traditional Recommendation systems are limited as they only consider certain metrics, such as product purchase details, product category. Recommendation systems are yet to gain popularity in the medical field. These days most patients are unable to figure out the medication that works in healing them in the best way possible, hence they turn to review websites in order to obtain a second opinion on the prescribed medication. In this work, we have developed a smart recommendation system for off-the Shelf Medical Drugs using machine learning and data analytics based on patient feedback. The patient feedback is unstructured data which is processed using data analytic tools. After which machine learning is used to recommend the best fit and compare the drugs. In this work, we predict the impact of a drug/ medicine on the patient to whom the medication was prescribed, using data mining techniques. Firstly, we detect the user's polarity (positive/ negative/neutral) based on the patient feedback for a certain drug using sentiment analysis and opinion mining following which we use machine learning algorithms to track sentiment variation and to make a recommendation based on user polarity

1. Introduction

These days, a large amount of personal information is available online in the form of online textual reviews. A large percentage of people make decisions based on these reviews that are available online. For instance, a patient may consider looking up at online reviews before using any medicine. From this, we can draw a conclusion that a positive review by the previous users can be used to suggest and recommend the medicine to potential users. This leads us to a point where it is of utmost importance that the relation between the reviews and reviewers in social networks be considered as a primary aspect in web mining, machine learning and natural language processing.

Natural Language Processing plays an important role in recommendation systems. Sentiment Analysis is used to understand the mood of the users of a particular product or issue. Sentiment Analysis is also known by the name Opinion Mining. The public usually expresses its opinions in blog posts, comments, reviews, tweets etc. We need to build a system that collects and examines the opinions and finds the variations in sentiment tracking. Opinion mining and Sentiment Analysis have several applications. When a company wants to launch a new product or if it wants to know which version of its products or services is preferred amongst the public, or to find which demographics prefer which product, Sentiment Analysis is very useful.

However, Sentiment Analysis has many challenges. Each person's way of expressing opinions is different. The major challenge is to identify if the opinion word is positive or negative. Some words maybe considered positive in certain situations and negative in other situations. For instance, consider good as a positive opinion. "This book is good" is a positive

opinion while "This book is not good" is a negative opinion. The fact that most of the text processing relies on the factor that a slight difference between two words does not change the overall meaning of the sentence makes it difficult to identify the polarity of the sentiment of the opinion holder. It is easier to identify if a review is positive or negative by examining one sentence at a time. However, these days people prefer informal mediums such as blogs, tweets or any other social networking website for that matter. They combine different types of opinions in the same sentence which can be comprehended by the human brain but is difficult for the computer to comprehend and determine the sentiment.

Generally, opinions may be expressed on various products e.g., a product, a service, a topic, an institution, a law etc. The term object is used to denote the entity on which an opinion has been expressed. Each object had a set of components and attributes that define its existence. The components may be further sub-divided into sub-components and each of these sub-components may have a separate set of attributes. The object is associated with a pair, $O:(T,A)$, where T is the hierarchy or taxonomy of components and sub-components of O, and A is a set of attributes of O. Each component has its own set of sub components and attributes.

Considering hierarchy, the object is the root node. The non-root nodes are the components and the sub-components. Each of these non-root nodes has a set of attributes. Generally, an opinion may be expressed on any node or on any attribute. This hierarchical representation is too complex for an ordinary user to understand. Hence, this hierarchical representation or tree is "flattened". The word which is used to represent both components and attributes is "features". The object which is the root of the tree is also a feature. An opinion on an object can be expressed in various platforms such as blogs, review websites etc. Opinions are expressed in the form of sequences of sentences. This sequence of sentence is known as Opinion passage of a feature. This Opinion passage of a

feature is a sequence of sentences that expresses if the opinion is positive or negative. However, there are various instances where in a single sentence may express opinions on more than one feature. Opinion holders also play a major role because they determine the individual or organization that holds the opinion.

The sequence of sentences are evaluated to find the polarity of the user. These evaluative texts at the document level or sentence level are not sufficient to find the user polarity. A document with a dominant amount of positive opinions does not necessarily guarantee that the opinion holder has positive opinions about all the aspects of the objects. Likewise, a document with more negative opinions does not necessarily mean that the opinion holder dislikes every aspect of the object. In most cases, these evaluative documents contain both positive and negative opinions and therefore it is important to find the general sentiment of the opinion holder. In order to find out the general sentiment of the opinion holder, we need to consider the feature level of the object. To do this, three key mining tasks are of utmost importance. They are

- Object Feature Identification
- User Polarity Determination and Sentiment Variation Tracking
- Synonyms Grouping

Object Feature Identification finds the features of an object on which opinions are expressed. For example “The color of the book is red”, In this sentence the color is the feature. Features are found by this technique by finding frequent nouns and phrases. Determination of user opinion orientation is used to find if the opinions on the features are positive, negative or neutral. For example “The texture of the cloth is good” is positive, In this sentence, the opinion on “texture” which is the feature of the object “cloth” is positive. In some documents, the same objects features may be expressed with different words. Grouping synonyms groups those phrases together to make the determination of user polarity and sentiment variation tracking easier.

The availability and popularity of opinion rich resources is growing day by day. Therefore it is very important to understand that the various challenges and opportunities that arise with regard to this. Sentiment Analysis and Opinion mining have gained a lot of importance especially when it comes to modeling and building recommendation systems. In this work we perform opinion mining on patient’s utility about medicines. On the basis of this we design and develop a recommendation system for off-the shelf medicines.

2. Literature Survey

Sanjiv R. Das Et al created a methodology for mining sentiments of small investors from stock message boards. The advancement in technology has made it possible to access the continuous flow of the public opinions from the World Wide Web. The stock market is largely being influenced by the Sentiment posts on the stock message boards. A method to extract the small investor’s sentiment to create an index which is not in existence was proposed in this paper along with the illustration of the applications of this methodology. The methods proposed have helped us to decide on the algorithms that we would like to use for the testing of our dataset.

Erwan Le Martelot et al have explored the use of stability optimization using greedy algorithms in the field of multi scale community detection. According to this paper, Networks can be used to represent real time systems. Analysis of such networks proves to be beneficial in gathering information about the organization of the original system. A new measure was introduced for partition quality. The new measure is stability. Several variations and heuristics as well as an application to overlapping communities has been presented with optimizing algorithms.

Jayashri Khairnar et al proposed the use of Support Vector for

classification of the sentiments. It is important to classify and organize opinions, suggestions, feedback and views due to the huge amount of data that is available of the Web. Therefore, Opinion Mining and Sentiment Analysis play an important role in identifying the opinions of the users as positive, negative and neutral. The proposed theory uses Maximum Entropy, Support Vector Machine and Naïve Bayes for Sentiment Analysis. Sentiment classification accuracy is also considered. The methods proposed helped us identify the algorithms that work the best for Opinion mining and Sentiment Analysis.

B. Sampath Kumar Et al have discussed the various methods of Opinion Mining. The proposed method was to process the data from the search results obtained when a search is made for the given item. Rating of the product can only be made by analyzing the reviews given by the customers.

Altug Akay Et al proposed a two-step framework to find the sentiments of the user along with the uses and side effects of cancer treatment Word frequency is mapped by using a self organizing map and a network-based approach and modeling the user sentiments from the forum interactions and employing a network partitioning method to optimize the quality measures stability. Identification of influential users and determination of consumer opinion would be done more efficiently.

3. Background Information

Naïve Bayes Classifier

The Naïve Bayes Classifier algorithm, mostly used in text classification, has higher success rates when compared to other algorithms. As a result, it is widely used for spam filtering and sentiment analysis.

Bayesian Network

Bayesian Network Algorithm is used to evaluate the prediction using different performance metrics on the trained data set.

4. Proposed Methodology

The proposed system has the following tasks:

- Spam Detection and Preprocessing of Data Set

In this work, we have used the patient feedback (sentiments) on medical drugs, which is unstructured data expressed mostly in the internet language. First, spam detection is done in order to avoid redundancy and enhance the reliability of the data set. After which removal of stop words and emoticons is done so that only valuable information in the data set is taken into consideration.

- Sentiment Analysis

In this work, we have used Naïve Bayes classifier and collaborative filtering for spam filtering and sentiment analysis of patient feedback, based on which the user polarity is detected.

The Naïve Bayes model involves an assumption for enhancing a simple conditional independence. In a given class (Positive/Neutral/Negative), the words in the class are conditionally independent of each other. This assumption enables the application of really fast classification algorithms for the problem in hand, without affecting the accuracy in text classification. Narayanan V Et al explore and discuss the accuracy of Naïve Bayes classifier for sentiment analysis in their 2013 paper [8].

In our work, the probability that a word belongs to a particular class is given by the following equation.

$$P(e_i | g) = \frac{\text{Count of words } e_i \text{ in all documents of class } g}{\text{Number of words in all documents of class } g} \quad (1)$$

Where,

g- A given class of documents (Positive/negative/neutral)

i- A word in all documents of a class g

And based on naïve Bayes algorithm, the probability that a document D belongs to a class g is given by:

$$P(g_i | D) = \frac{p(D | g_i) * p(g_i)}{P(D)} \tag{2}$$

Where,

g_i – Is the class

D- A given document of words and,

$$P(g_i | D) = \frac{(\prod p(e_i | g_i)) * P(g_i)}{P(D)} \tag{3}$$

Where,

g- A given class

i -Words in a given document D

D- A document D that belongs to a class g

• Sentiment Variation Tracking

The next task in our work is to track the variations in user sentiments.

Bayesian Network Algorithm is used to evaluate the prediction using different performance metrics on the trained data set. After which, A recommendation is made for a certain drug based on the user sentiment polarity.

We have created a Bayesian Network that will allow us to determine the exact numerical implications of the user feedback on the drug’s performance. In the proposed system, the Bayesian network will contain two nodes representing Random Variables, the nodes are:

User Polarity (Positive/negative/neutral)

Final Recommendation

• Final Recommendation

The Final recommendation system rates the medical drugs and recommends the best fit based on data records and user polarity. The Final recommendation system also takes into account the number of positive, negative and neutral reviews and makes recommendation on the basis of this count

The work in total maybe represented as in Fig 1

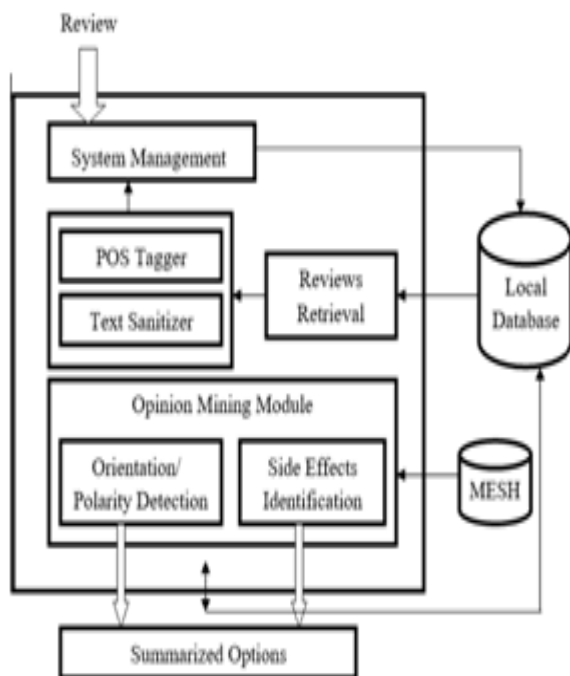


Fig. 1

In this work, we first preprocess the data sets , wherein we perform spam detection to avoid redundancy and remove stop words and emoticons from the patient feedback. After which, we use naïve Bayes classifier algorithm for sentiment analysis and sentiment variation tracking. Finally, we use Bayesian Network algorithm to make a final recommendation based on user polarity.

5. Result Analysis

The proposed system was tested with patient reviews. A comparison of the performance of the proposed system and the existing one is presented in terms of precision and recall in Fig2.

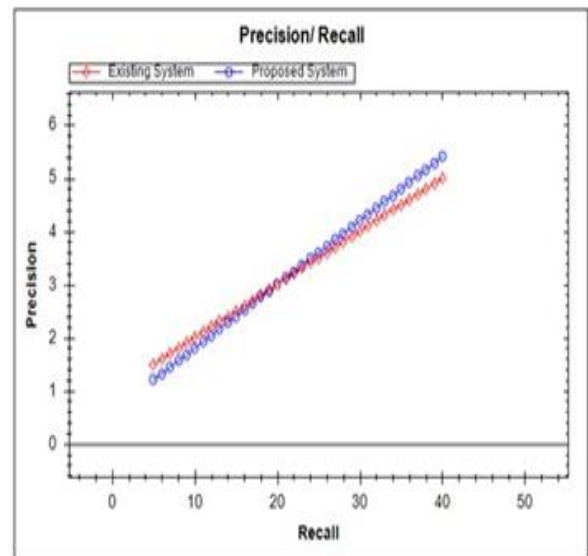


Fig. 2

There is a marginal difference between the existing system and proposed system. The slope of proposed system makes an improved balance between precision and recall, as a result of which more relevant reviews e taken into consideration.

6. Conclusion

In this Paper, a smart recommendation system is proposed by mining patient feedback on medical drugs and sentiment information from Patient reviews. The proposed system employs data analytic tools to preprocess unstructured data and train the datasets. After which, sentiment analysis is used to detect user polarity and track variations in patient sentiments. Based on user polarity, a final recommendation of the best fit is made. The proposed system uses techniques to implement feature based opinion mining. The main aim is to help patients seeking for a second opinion on a prescribed medication. The proposed system uses feedback from patients who have already taken the same prescribed medication and implements data analytics to track the efficiency and effectiveness of the medication. Hence, this smart recommendation system will serve as a means of verifying the possible effects and effectiveness of a medication. There is a wide scope for future improvement in the proposed recommendation system. In the future work, the recommendation system can be enhanced by considering other metrics such as cost and healing time to make better recommendations and enable highly precise data analysis by upgrading the user feedback sentiment dictionaries.

7. Acknowledgements

The authors would like to thank their guide/ mentor Mrs. A.Meenapriyadarsini, Assistant professor and Faculty of Computer Science Engineering, SRM University and Mr. S.Prabakaran, Assistant professor (Sr. G), Faculty of Computer Science Engineering, SRM University for guiding and encouraging them throughout the project.

References

- [1] Xiaojiang Lei, Xueming Qian, Member, IEEE, and Guoshuai Zhao, "Rating Prediction Based on Social Sentiment from Textual Reviews", *IEEE transactions on multimedia*, Vol. 18, No. 9, September 2016.
- [2] A. Ng, A. Zheng, and M. Jordan, "Stable algorithms for link analysis," in *Proc. SIGIR Conf. Inform. Retrieval*, New Orleans, Louisiana LO, USA, 2001, pp. 258–266.
- [3] B. Taskar, M. Wong, P. Abbeel, and D. Koller, "Link prediction in relational data," in *Proc. Adv. Neural Inform. Process. Syst.*, Vancouver, B.C. Canada, 2003.
- [4] C. Corley, D. Cook, A. Mikler, and K. Singh, "Text and structural data mining of influenza mentions in web and social media," *Int. J. Environ. Res. Public Health*, vol. 7, pp. 596–615, Feb. 2010.
- [5] D. Liben-Nowell and J. M. Kleinberg, "The link prediction problem for social networks," *J. Am. Soc. Inform. Sci. Technol.*, vol. 57, pp. 556–559, May 2007.
- [6] L. Dunbrack, "Pharma 2.0 – social media and pharmaceutical sales and marketing," in *Proc. Health Ind. Insights*, 2010, p. 7.
- [7] Rennie, J.D., et al.: Tackling the poor assumptions of naive bayes text classifiers. In: *Machine Learning-International Workshop then Conference*, vol. 20(2) (2003).
- [8] Vivek Narayana, Ishan Arora, Arjun Bhatia, "Fast and Accurate Sentiment Classification Using an Enhanced Naïve Bayes Model", *International Conference on Intelligent Data Engineering and Automated Learning*.
- [9] J R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Proc. NIPS*, 2007, pp. 1257–1264.
- [10] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in *Proc. 18th ACM SIGKDD Int. Conf. KDD*, Aug. 2012, pp. 1267–1275.
- [11] M. Jiang et al., "Social contextual recommendation," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage*, 2012, pp. 45–54.
- [12] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proc. ACM Conf. Recommender Syst.*, 2010, pp. 135–142.
- [13] Z. Fu et al., "Achieving efficient cloud search services: Multi-keyword ranked search over encrypted cloud data supporting parallel computing," *IEICE Trans. Commun.*, vol. 98, no. 1, pp. 190–200, 2015.
- [14] G. Ganu, N. Elhadad, and A. Marian, "Beyond the stars: Improving rating predictions using Review text content," in *Proc. 12th Int. Workshop Web Databases*, 2009, pp. 1–6.
- [15] J. Xu, X. Zheng, and W. Ding, "Personalized recommendation based on reviews and ratings alleviating the sparsity problem of collaborative filtering," in *Proc. IEEE 9th Int. Conf. e-Bus. Eng.*, Sep. 2012, pp. 9–16.
- [16] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized recommendation combining user interest and social circle," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp. 1763–1777, Jul. 2014.
- [17] S. Li et al., "Sentiment classification and polarity shifting," in *Proc. 23rd Int. Conf. Comput. Linguistics*, 2010, pp. 635–643.
- [18] B. Sun and V. Ng, "Analyzing sentimental influence of posts on social networks," in *Proc. IEEE 18th Int. Conf. Comput. Supported Cooperative Work Des*, May 2014, pp. 546–551.
- [19] G. Zhao, X. Qian, and X. Xie, "User-service rating prediction by exploring social users' rating behaviors," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 496–506, Mar. 2016.
- [20] D. Vilares, M. A. Alonso, and C. Gomezrodríguez, "A syntactic approach for opinion mining on Spanish reviews," *Natural Lang. Eng.*, vol. 21, no. 1, pp. 1–25, 2014 [21] Sanjiv R. Das and Mike Y. Chen, "Yahoo! For Amazon reference: Sentiment Extraction from Small Talk on the Web, May 2, 2004.
- [21] Erwan Le Martelot and Chris Hanki, "Multi scale Community detection using stability as optimization criterion in a Greedy Algorithm", Imperial College of London. Department of Computing
- [22] Jayashri Khairnar and Mayura Khairnar, "Machine Learning Algorithms for Opinion Mining and Sentiment Classification", Department of Computing, Pune University, MIT Academy of Engineering, Pune.
- [23] B. Sampath Kumar and Dr D. Bhanu Shree Reddy, "An Analysis on Opinion Mining Techniques and Tools", VIT Business School, VIT University.
- [24] Altug Akay School of Technology and Health, Royal Institute of Technology, Stockholm, Sweden Andrei Dragomir School of Technology and Health, Royal Institute of Technology, Stockholm, Sweden Björn-Erik Erlandsson Department of Biomedical engineering, University of Houston, Houston, TX, USA Network based modeling and intelligent data mining of socialmedia for improvingcare.
- [25] T. Padmapriya and V. Saminadan, "Priority based fair resource allocation and Admission Control Technique for Multi-user Multi-class downlink Traffic in LTE-Advanced Networks", *International Journal of Advanced Research*, vol.5, no.1, pp.1633-1641, January 2017.
- [26] M. Rajesh, Manikanthan, "ANNOYED REALM OUTLOOK TAXONOMY USING TWIN TRANSFER LEARNING", *International Journal of Pure and Applied Mathematics*, ISSN NO:1314-3395, Vol-116, No. 21, Oct 2017.
- [27] S.V. Manikanthan, T. Padmapriya "An enhanced distributed evolved node-b architecture in 5G tele-communications network" *International Journal of Engineering & Technology (UAE)*, Vol 7 Issues No (2.8) (2018) 248-254. March 2018