

A Survey on Opinion Effects-Opinion based Learning Model in Medical Sector

Sagar B. S¹ Arpita Kulkarni² Chaithra³ Jyothi⁴ Natesh⁵

^{1,2,3,4,5}Department of Computer Science and Engineering

^{1,2,3,4,5}Vidyavardhaka College of Engineering, Mysuru, Karnataka, India

Abstract— The online healthcare community offers a wide range of detailed medical information to physicians, system administrators and patients. In this article, we collect real-time health publications from popular websites, where patients can demonstrate their perspectives, experiences, and use of medications. To this end, we intend to summarize all user publications for each drug and summarize the medical organization and patient population. Subsequently, we classify all users according to their "mental state" and, in addition, we analyze consumer information through which we can find useful "Relations" for drugs with drug symptoms. "This can be done using the rules of the Association Rule."

Keywords: Medical Sector, Learning Model

I. INTRODUCTION

To understand the role of health, we proposed the use of certain separate operations, such as: summary, association of association rules and mood analysis for data obtained from the health site.

Summary: It is intended to download details of the source, extract the content and show the user useful content in a modified and satisfactory way for the user's application [1].

This is very important in many NLP applications, such as quality analysis, information retrieval, etc.

There are two types of abstracts: 1. Extract the text, that is, sentences and words are used repeatedly. 2. Summary that includes the rejuvenation of the content downloaded from the texts.

Association rules mining: this is a popular and well known activity in data mining. It is used to discover the relationships between variables in a large data set. The rules generated by extracting the association rules will have elements disconnected in series 2

LHS => RHS. This means that HRH will be more likely when the LHS series occurs [2].

Data extracted obtained by applying association rules

- 1) Age group -> disease
- 2) symptoms -> disease
- 3) medicine-> disease
- 4) disease-> disease
- 5) disease-> drugs

Feeling analysis: the task is to find the feelings of the text. Feelings can take many forms: emotions and attitudes towards the organization / being, people's opinions. This entity can represent events / entities, people. These topics are covered by notifications. Sentiment analysis is considered a classification process. Classification can be carried out at the level of prayer, document and appearance. When analyzing the feelings, all important attributes are selected in the text, so the classification process is carried out using the appropriate classifier.

A. Existing Methods

Health communities simply record a real-time health report, in which patients express their opinions, including their experiences and their side effects with the medications they consume.

B. Drawbacks

Just keep the health position

There is no summary

It is not possible to extract useful information

Reduction of user satisfaction

Stores large amounts of data that are difficult to analyze.

II. RELATED WORKS

Harsha Ranjan, sumit agarwal, amit prakash and Sujana Kumar Seha have proposed a model which labels the important keywords which are used in medical discussions. This model has 2 approaches namely, syntactic approach and semantic approach. Syntactic approach has 3 methods (word extraction based on its frequency, extraction based on topic modeling). Semantic approach has 2 methods (resolution of the pronoun and labeling using semantic role). Dataset is obtained from doctors.ndtv.com. Combining both syntactic and semantic model system had achieved 83.38%, syntactic had achieved 59% and semantic had achieved 71% [3].

Vinod L. Mane, Suja S. Panicker and Vidya B. Patil have proposed a model which specifies the association between drug, disease and symptoms. This model uses apriori algorithm to find relation between parameters in dataset. Dataset is obtained from healthboard.com. Model has 2 steps (extracting keywords and association between the parameters). Association is done in 2 steps (Association rule generation and interesting rule selection). Drawback in this model is they are not concentrating on symbolic comments and spelling mistakes [4].

Yi Chen and Yunzhong Liu have proposed a model which uses the thread based search which helps the users by providing the details of medicine related to particular disease and adverse drug reaction discovery based on the analysis of co -occurrence. Dataset is obtained from Patientlikeme.com or medhelp.org, Drawback in this model is that it fails to determine whether given disease is a symptom or adverse effect because of the significant overlap between vocabulary [5].

AlokRanjan Pal and DigantaSaha have proposed a model which uses LESK algorithm for the summarizing text. LESK uses semantic analysis to extract relevant sentences or group of keywords. This model uses WorldNet as a semantic dictionary. LESK algorithm will solve the problem of ambiguity. This model uses unsupervised learning. Time complexity of the algorithms used here is $O(n^3)$, time complexity is one of main drawbacks of this model. But unsupervised approach does job better than rule approach [6].

Rafael Ferreira, Frederico Freitas and Luciano de Souza Cabral have proposed the model which recommends and brings trial proof that the adequacy of sentence scoring strategies for programmed extractive content synopsis calculations relies upon the sort of content one needs to outline, the length of records, the sort of language utilized, and their structure. This model has 3 approaches. Word based scoring, graph based scoring and sentence based scoring. Word based scoring uses word frequency, tf/idf, proper nouns and lexical similarity as scoring methods. Graph based scoring uses bushy path or aggregate similarity as scoring methods. Sentence based scoring uses cue phrases, centrality, length and inclusion of sentence as scoring methods [7].

This link will give you brief introduction about sentiment analysis and how sentiment analysis divides the group of sentences into positive and negative sentence. Basically it has 4 steps. Tokenization (splitting of data taken place), cleaning of data (removing special characters), Removing stop words (removing the words like 'the', 'was', 'is', etc.) and classification (It will classify the words in a sentence into positive or negative or negative or neutral). Lexicons are used as dictionary here. Final output will be polarity and subjectivity of the sentence [8].

Lakshmi K.S and G. Santhosh Kumar They proposed a model to which they pointed, extracting all the useful information from medical transcription transcripts. They used a set of semi-structured and unstructured transcripts, in this they used data mining and NLP. The main objectives of the data mining is to extract all the useful content from existing data and the extracted content is useful for later use. The association rule, is a very popular and sought after method to create relationships between variables present in a large data set. It consists of 2 steps: 1. Generation of the Association rules 2. Interesting selection of rules. For structured XML documents to generate the association rules, they used the FP Growth Algorithm and the Apriori algorithm. Weka is used to implement the Apriori algorithm. The entry to generate the association rule is taken from the XML documents. For rules the elevation value is calculated for the selection of interesting rules. The rules obtained are valid and clinically proven [9].

Subhabrata Mukherjee, Gerhard Weikum and Cristian Danescu-Niculescu-Mizil, They proposed a model in which they focused mainly on the reliability of statements about side effects in the health community. They proposed a graphic model to learn together the interactions between the credibility of the statements, the reliability and the language used. To know the reliability of user statements, they extract all the side effects of each drug, in which they can capture user interactions, emotional language and the style they used and user properties. This model was able to identify unknown side effects of all drugs. Compared to baselines (57.6%), this model, which uses the supervisory classifier and the broad set of the features for users, language and publication, achieves average improvement of 11.4% by increasing accuracy [10].

Jesmin Nahar, Tasadduq Imama, Kevin S. Tickle, Yi-Ping Phoebe Chen they had proposed a model in which they had widely used the association rule. They used 2 separate LHS and RHS sets, where the RHS sets is more likely to appear whenever, the LHS sets appears. Confidence and support are the two measures of the intersection pattern that

will reflect the functionality, certainty of the pattern. Mining association rule is a very useful method, especially in the medical field, to know the relationships between variables. This technique has been explored in 3 types: 1: Apriori, 2. Predictive Apriori 3. Tertius. They focused primarily on computational intelligence, particularly the association rule, Data Mining classifiers, to identify all key factors underlying a given disease, as well as the gender diversity considered. [11].

Kostas Fragos¹, Yannis Maistros¹, Christos Skourlas He has suggested a model that uses Internet vocabulary and databases to decode the meanings of words. In the statistical approach, there is a problem with the "data shortage", to decode words of several words and overcome this problem, they use two important elements: the iteration to find similarities and transit these similarities, which allows us to obtain similarities of order higher. The disadvantage is that text search in this model refers to the problem of finding relevant documents for specific questions in a text [12].

Sunil Kumar and Maninder Singh The health industry has proposed an important data analysis model: Impacts, applications and tools, this article has provided an in-depth report and a brief description of the key data of general and health systems that play an important role in Health information which greatly influences the health services system and the four key data of IHealthcare V. They also suggest using a conceptual architecture to address big data health problems using Hadoop-based big data usage terminology generated by different levels. of medical data and developing methods to analyze the data. and get answers to medical questions. The effective treatment of certain patients can lead to a combination of large amounts of data and health analysis, since it allows prescribing the right medication for everyone, not for most people. [13].

III. CONCLUSION

An important research area is the exploration of the archives of the users of the health areas for the dissemination of information. Our work will encourage patients to establish links between different diseases, medications and symptoms. This will also help doctors to discover the side effects of different medications so they can prescribe the best / different medications for other patients with the same disease. Even pharmaceutical companies are making a profit because individual addicts are classified as poor, normal and happy. This category can be an indirect entry for companies if they have to decide which medication to use and if the company must produce alternative medicine for certain diseases, etc. Fraternities and pharmaceutical companies

REFERENCES

- [1] Jayashree R, Srikanta Murthy K, Basavaraj .S. Anami, "Categorized Text Document Summarization in the Kannada Language by Sentence Ranking", 12th International Conference on Intelligent Systems Design and Applications (ISDA), pp 776-781, 2012.
- [2] Jesmin Nahar, Tasadduq Imam, Kevin S. Tickle, Yi-Ping Phoebe Chen, "Association rule mining to detect factors which contribute to heart disease in males and females",

- J. Nahar et al. / Expert Systems with Applications 40 (2013) 1086–1093, Elsevier, 2012.
- [3] Harsha Ranjan, Sumit Agarwal, Amit Prakash and Sujana Kumar Saha. 2017. Automatic Labelling of Important Terms and Phrases from Medical Discussions. In Conference on Information and Communication Technology (CICT 2017), ABV-IITM Gwalior, Nov 3-5, 2017.
- [4] Vinod L. Mane, Suja S. Panicker, Vidya B. Patil. "Knowledge discover from user health posts", 2015 Ieee sponsored 9th international Conference on intelligent system and control (ISCO).
- [5] Yi Chen, Yunzhong Liu, "Connecting the Dots: Knowledge Discovery in Online Healthcare Forums", ICEC'14 August 05 - 06 2014, ACM.
- [6] AlokRanjan Pal, DigantaSaha, "AnApproach to Automatic Text Summarization using WordNet", IEEE International Advance Computing Conference (IACC),2014.
- [7] Rafael Ferreira, FredericoFreitas, Luciano de Souza Cabral, Rafael DueireLins, Rinaldo Lima, Gabriel Franca, Steven J. Simske, and Luciano Favaro, "A Context Based Text Summarization System",11th IAPR International Workshop on Document Analysis Systems,pp 66-70,2014.
- [8] [https://www.youtube.com/watch?v=O_B7XLfx0ic&t=318\\$](https://www.youtube.com/watch?v=O_B7XLfx0ic&t=318$).
- [9] Lakshmi K.S and G. Santhosh Kumar "Association Rule Extraction from Medical Transcripts of Diabetic Patients"
- [10] Subhabrata Mukherjee, Gerhard Weikum and Cristian Danescu-Niculescu-Mizil "People on Drugs: Credibility of User Statements in Health Communities"
- [11] JesminNahar,TasadduqImama, Kevin S. Tickle ,Yi-Ping Phoebe Chen" Association rule mining to detect factors which contribute to heart diseasein males and females
- [12]Kostas Fragos¹, Yannis Maistros¹, Christos Skourlas Using WordNet Lexical Database and Internet to Disambiguate Word Senses.
- [13]Sunil Kumar and Maninder Singh 'Big Data Analytics for Healthcare Industry:Impact, Applications, and Tools'ISSN[2018].