



## SENTIMENT BASED DOCTOR AND HOSPITAL RATING SYSTEM

**Mr. Biju  
Balakrishnan**

Associate Professor Department of Computer Science and Engineering, J.C.T College of Engineering and Technology, Pichanur, Coimbatore, TamilNadu, India

**ABSTRACT** Online health communities continue to offer huge variety of medical information useful for medical practitioners, system administrators and patients alike. In this work we collect real time health posts from reputed websites & social medias where patients express their views, including their hospital and doctor experiences and side-effects on drugs used by them. We propose to perform Summarization of user posts per drug, and come out with useful conclusions for medical fraternity as well as patient community at a glance. Further, we propose to classify the users based on their 'emotional state of mind', and the user can find the best doctor an hospital in nearest locality with the help of GPS. Also, we shall perform knowledge discovery from user posts, whereby useful informations about the hospital, doctor, drug etc are done by Association Rule Mining.

**KEYWORDS :** Association rule mining, Classification, Keyword extraction, Knowledge discovery, Summarization

### 1)INTRODUCTION

With the enormous increase in web, electronic information is also increasing in huge amount which, although good with respect to Information Age, creates overhead of time and space. Also understandability of information and consequent knowledge continue to be big challenges. For knowledge mining of the health posts, we propose to apply different important operations like - Association Rule Mining, Summarization and Sentiment Analysis on data obtained from the health social medias and reputed websites.

Summarization is very important in different NLP applications like Information Retrieval, Quality Analysis, Text Comprehension etc. Commonly there are two types of summaries. First one is Extract in which contents from text i.e. words and sentences are reused. Second one is Abstract which includes regeneration of extracted contents [2]. Association rule mining is a popular and widely-known data mining task. It is used to find out interesting relations between variables in large database. Rules generated by association have two disjoint set of items having form LHS (Left Hand Side) => RHS (Right Hand Side). The rule says that RHS is likely to occur whenever the LHS set occurs [3]. Extraction of association rules includes two steps[4]: 1. Association Rule generation 2. Interesting Rule Selection

Sentiment Analysis (SA) or Opinion Mining (OM) is task of finding sentiments from text. These sentiments may take different forms like – opinions from people, attitudes and emotions toward an entity. The entity can represent individuals, events or topics. These topics are most likely to be covered by reviews. WalaaMedhat considered Sentiment Analysis as a classification process. Classification levels considered were - document level, sentence level and aspect level [5]. While doing SA first the important features are selected from text then classification is done using appropriate classifier. We are considering reviews from health posts and in our case represented entity is drug. So our classification falls in aspect level

### I.PROPOSED ALGORITHM

#### A.Keyword Extraction

In this module, input data is taken as user posts obtained from healthboards.com. Each keyword from post is assigned to particular UMLS (Unified Medical Language System) category like organic chemical, sign, symptom, disease, feeling etc. Then keywords having the category disease, drug or symptoms are extracted for further processing

example:

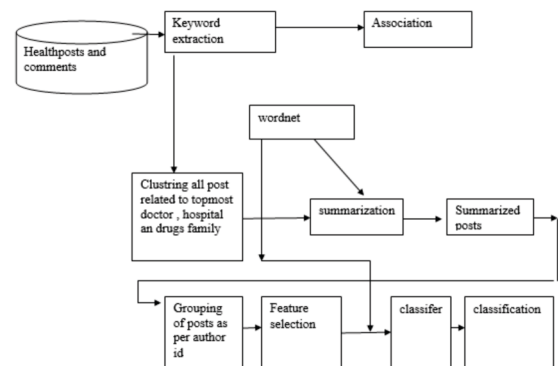
3 102343 Am suffering from cold and fever since few months. just i took calpolbt am not getting the relief from ma disease.

Fig.1 shows sample post for drug calpol. In this figure '3' is author id, '102343' is document id and rest of the part is actual post. Fig.2 shows assignment of UMLS categories to each keyword (Keyword: UMLS category).

UMLS categories:

suffering: **finding** cold: **Symptom or disease** fever: **finding** few:

quantitative concept months: temporal concept calpol: organic chemical



#### B.Association

From extracted keywords received in above module we propose to determine the different types of association. These associations could be among disease, drug and symptoms. We shall use Apriori algorithm for this purpose.

#### C.Summarization

We are concentrating on summarization of topmost drug family only. So first we will cluster together all the posts related to topmost drug family. Then using Lesk based summarization algorithm [1] we will be generating summary. We are using Wordnetdictionary [13] to detect correct sense of word. It would help to generate better summary results.

#### D.Sentimental Analysis (SA)

Text not only provides informative content but also attitudinal information, including possible emotional states of the user who posted his views. For SA, we are collecting all the posts per author at one place. To avoid unnecessary preprocessing we are taking output of summarization as input for this module. We are using bag-of word approach and tf-idf feature selection then using classifier we propose to classify user into different classes like normal, depressed and satisfied.

#### II.FUTURE SCOPE

Social media posts contain a lot of errors or spelling mistakes. We are not considering spelling mistakes and their correction. So this could be further improvement. Posts in social networking may also contain symbolic expressions, which are not considered in this work.

#### III.CONCLUSION AND FUTURE WORK

Analyzing user posts and comments from health communities and social media for knowledge discovery is an interesting area in research. This work will help patients to find out association among different hospital an doctors near his/her location ,drugs, diseases and symptoms. It will help doctors to find out side-effects of different drugs so they can prescribe better drugs to other patients with similar disease.

Pharmaceutical companies will be also benefited as we are classifying users of particular drug into different classes like normal, depressed and satisfied. This will be indirect input to companies to decide which drug is popular, whether to produce alternate drug to this etc. Thus our work shall equally benefit all three parties—medical fraternity, patient community and pharmaceutical companies.

## REFERENCES

- [1] JayashreeR,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] AlokRanjan Pal, DigantaSaha, "An Approach to Automatic Text Summarization using WordNet", IEEE International Advance Computing Conference (IACC), 2014.
- [3] JesminNahar, 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.
- [4] Lakshmi K.S, G. Santhosh Kumar, "Association Rule Extraction from Medical Transcripts of Diabetic Patients", IEEE, 2014.
- [5] WalaaMedhat, Ahmed Hassan, HodaKorashy, "Sentiment analysis algorithms and applications: A survey", In press, Elsevier, 2014.
- [6] 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.
- [7] C. Lakshmi Devasenal and M. Hemalatha, "Automatic Text Categorization and Summarization using Rule Reduction", IEEE- International Conference On Advances In Engineering, Science And Management (ICAESM -2012), pp 594-598, 2012.
- [8] Sara Keretma, CheePeng Lim, Doug Creighton, "A Hybrid Model for Named Entity Recognition Using Unstructured Medical Text", Proc.Of the 2014 9th International Conference on System of Systems Engineering (SOSE), Adelaide, Australia June 9-13, pp 85-90, 2014.