WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Research Paper on A Smart Medical System for Drug Analysis Using ML Algorithms

Prof. Manjunath N[1], Ananya S [2], Anushree S[3], Chandana S[4], Kavya S P[5]

[1]Assistant Professor, Department of Computer Science and Engineering
[2, 3 ,4,5] Student, Department of Computer Science and Engineering
Vidya Vikas Institute of Engineering and Technology, Mysuru, Karnataka, India

ABSTRACT –

Online health communities continue to offer huge variety of medical information useful for medical practitioners, system administrators and patients alike. In this system we collect real time health posts from reputed websites, where patients express their views, including their 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'. Also, we shall perform knowledge discovery from user posts, whereby useful `patterns' about the triad `drugs-symptoms-medicine' is done by Association Learning.

I. 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 understand-ability 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 forum site health-boards.com.

Summarization is defined as taking information from the source, extracting content from it, and presenting the most useful content to the user in a condensed form and in a manner suitable to the user's application needs [1]. Summarization is very important in different NLP applications like Information Retrieval, Quality Analysis, and 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 machine learning 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 islikely to occur whenever the LHS set occurs [3].Extraction of association rules includes two steps [4]:

1. Association Rule generation

2. Interesting Rule Selection

After the rules have been obtained, they are extracted and post processed. The extracted rules from the health boards data-set could take one or more of the following form-

1. Symptoms->disease

2. Disease->disease

3. Medicine->disease

4. Disease->medicines

5. Age group->disease.

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. Walaa Medhat 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.

II. METHODOLOGY

Step 1: In the 1 st step we collect the medical data (patient opinions[feedback] based on drugs). we referred sources like www.patientslikeme.com,www.healthboard.com and www.kaggle.com.

Step 2: Patient opinions are then preprocessed and irrelevant data removed and only relevant data extracted and inputted to the algorithms.

Step 3: Then we input the necessary things required for algorithms, initially preprocessed patients opinions are summarized using lesk based algorithm and output of lesk based algorithm is inputted to Eclat algorithm to discover the medical patterns (symptomsdiseases-drugs).

Step 4: Eclat algorithm will discover the medical patterns which shows the relationship between symptoms with symptoms, symptoms with disease, disease with disease and disease with drug.

Step 5: Medical patterns displayed on GUI (front end).

Step 6: Results of the data science algorithms analyzed and represented visually

Figures and Tables

Figures and tables must be centered in the column. Large figures and tables may span across both columns. Any table or figure that takes up more than 1 column width must be positioned either at the top or at the bottom of the page. Graphics must not use stipple fill patterns because they may not be reproduced properly. Please use only *SOLID FILL* colors which contrast well both on screen and on a black-and-white hardcopy, as shown in Fig. 1.

The project architecture incorporates advanced technologies with a central server, user interfaces, and crucial Medical Practices. These practices create datasets on symptoms, diseases, drugs, and sentiments. User interaction involves an Opinion Mining System and Patients' Prediction for analyzing health opinions and predicting trends. The system efficiently processes large health datasets, and the user interface includes home, about, and contact pages.

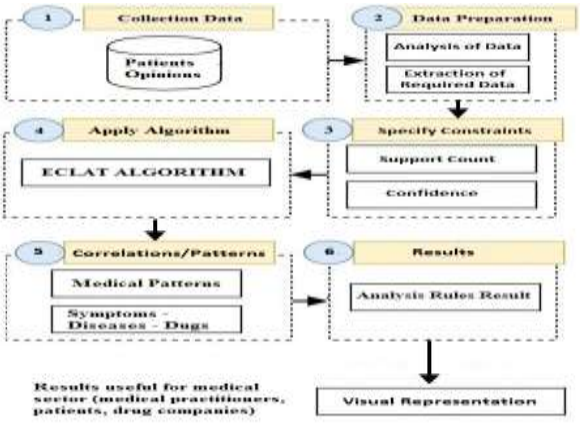


Fig 1 : System Architecture

Step 1: Data Collection: We are working on real time application, we build a new application which contains data servers (used to store data). Data collection means collecting data from different sources.

Step 2: Data Preparation: Here data from servers extracted and analyzed. Complete data extracted and analyzed where we remove irrelevant data and retain data required for processing. According to the project only symptoms , diseases and drugs are required to generate outputs.

Step 3: Specify Constraints

SUPPORT COUNT: The relationship between the total number of transaction containing that item (A) with the total number of transaction in data set.

CONFIDENCE: Confidence of item set defined as total number of transaction containing the item set to the total number of transaction containing LHS.

Step 4: Association Rules Mining (Eclat Algorithm): Association (or relation) is probably the better known and most familiar and straightforward data mining technique. Here, we make a simple correlation between two or more items, often of the same type to identify patterns.

For example, Market-basket analysis, where we track people's buying habits, we might identify that a customer always buys cream when they buy strawberries, and therefore suggest that the next time that they buy strawberries they might also want to buy cream.

We use eclat algorithm to process data and to find the patterns.

Eclat algorithm is selected because of the following reasons.

1. Quicker Results (takes less time for Prediction)

2. Works fine for small data set as well as Huge data set.

3. One scan of Database is Enough.

4.Works fine for multiple constraints.

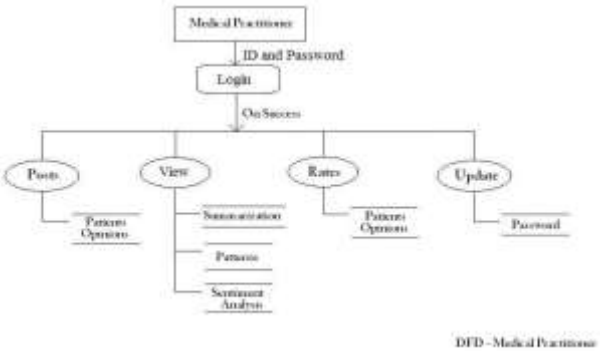
Step 5: Patterns Prediction: Here system predicts the relationship between symptoms, diseases and drugs.

III. FLOWCHART

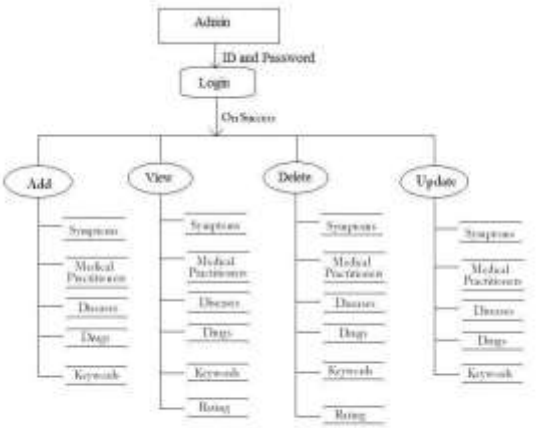
A data flow connects the output of an object or process to the input of another object or process. It represents the intermediate data values within the computation. It is draws as an arrow between the procedure and the consumer of the data value. The arrow is labeled with the description of the data, usually its name or type.

An actor is an active object that drives the data flow graph by producing or consuming values. Actors are attached to the inputs and the outputs of a dataflow graph. In sense, the actors lie on the boundary of the flow graph but terminate the flow of data as sources and sinks of data, and so are sometimes called terminators.

A data store is a passive object within a data flow diagram that stores data for later access. Unlike an actor, a data store does not generate any operations on its own but merely responds to requests to store and access data.



DFD - Medical Practitioner



DFD - Administrator

Fig 2 : Flowchart

The Fig 2 flowchart is designed to guide users through a personalized fitness and diet regimen, incorporating interactive elements and artificial intelligence for pose prediction and diet suggestions. It emphasizes the importance of correct exercise form and provides a structured approach to health and fitness goals

IV. CONCLUSION

In this work, we collect real time health posts from reputed websites, and perform data mining to determine the various possible associations from these posts and perform knowledge discovery from user posts and detect useful 'patterns' about groups like: disease to disease, disease to drug and drug to symptom. This is done using Association rules algorithm. This will help the doctors to find side-effects of different drugs and with this they can prescribe better drugs to other patients with similar disease. Pharmaceutical companies can the response of several drugs on people and will get a idea about which drug is popular and should be produced. This will also help the patients to know about the opinion of previous users, thus will be in a better position to decide which medicine should be taken for a particular disease and also improve awareness on various side-effects of drugs faced by other people.

V. 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, 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 Keretna, 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. 21

[9] SaeedMohajeri,AfsanehEsteki, Osmar R. Zaiane and DavoodRafiei, "Innovative Navigation of Health Discussion Forums based on Relationship Extraction and Medical Ontologies",IEEE International Conference on Bioinformatics and Biomedicine, pp 13-14, 2013.

[10] Yi Chen, Yunzhong Liu, "Connecting the Dots: Knowledge Discovery in Online Healthcare Forums", ICEC'14 August 05 - 06 2014, ACM.