IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9739

Opinion Based Learning Model In Medical Sector Using Machine Learning

Tejaswini N R[1], B M Bhavya[2]

PG Student, Dept of MCA, P.E.S College of Engineering Mandya, Karnataka, India[1]

Assistant professor, Dept of MCA, P.E.S College of Engineering Mandya, Karnataka, India[2]

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 patients and 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.

Keywords: Summarization, Sentimental Analysis ,Patient Opinions, Popularity Of Drugs.

I. INTRODUCTION

The 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 performed on patients opinions stored in database as data set and extracting content from patient opinions 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, 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 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

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- symptoms->disease, disease->disease,medicine->disease,disease->medicines,Age group->disease.

Sentiment Analysis is task of finding sentiments from text. These sentiments may take different forms like – opinions from people, attitudes and emotions toward an entity. 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.

II. LITERATURE SURVEY

Knowledge discovery from user health posts : Online health communities offer tremendous medical information that could be available to all. However as is the case with other data intensive applications, this information contains hidden patterns which if explored, analyzed and understood could be very useful for administrators , medical practitioner sand patients alike. There are websites where patients express their experiences or side-effects on drugs used. In this work we collect such real time health posts from reputed websites, and perform data mining to determine the various possible associations from these posts. Also, we shall perform knowledge discovery from user posts, whereby useful 'patterns' about groups like: disease to disease, disease to drug and drug to symptom are discovered.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9739

III. PROBLEM STATEMENT

In medical sector diagnosing a disease is a complex process and may not give you good results as it is manual process. Performing summarization of user posts per drug, and come out with useful conclusions for medical fraternity as well as patient community is a important factor in medical sector. For a disease, one(doctor) should give the proper treatment for the patient. As multiple drugs are available for the particular disease, there is need of identifying the popular drug. As symptoms are related to disease and diseases are related to drugs, there is a need for the system which discovers the relationship between symptoms-diseases-drugs. In the proposed system we are achieving this based on the patient opinion.

IV.PROPOSED SYSTEM

Proposed system collects real time health posts from reputed websites, where patients express their views, including their experiences and side-effects on drugs used by them. Proposed system perform Summarization of user posts per drug, and come out with useful conclusions for medical fraternity as well as patient community at a glance. Also, proposed system perform knowledge discovery from user posts, whereby useful `patterns' about the triad `drugs-symptoms-medicine' is done by Association Rule Mining.

System Architecture

The Application Manager or Admin will manage medical practitioners,word-net and database.The patients opinions are collected from reputed websites such as healthboard.com and all the information are stores in server and accessed through web browser and downloaded and store as trained data set to opinion mining system.Admin will register medical practitioners and generates unique id and password to medical practitioners ,with registration done by admin they cannot accessed the application.The admin views patient opinions of drugs and create and manage data sets, add keywords to database related to symptoms,disease and drugs and also have option to change password.
Medical practitioners will survey patients and uploads patients opinions with patients information,ratings of drugs to database.The main objective of project are summarization,medical patterns and sentimental analysis. Summarization is performed on data set and it analysis each patients opinions and output of summarization is inputted to medical patterns .medical patterns discovers five different patterns and identifies relationship between symptoms ,disease and drugs.sentimental analysis is performed on patients opinions it shows satisfaction of drugs posted by patients and positive and negative opinions of patients related to drugs and it help medical practitioners to prescribe better drugs to patients and also help pharmacy companies to produce popular drugs.The visitors open application and view patients opinions and popularity and satisfaction of drugs and side effects of drugs posted by patients.

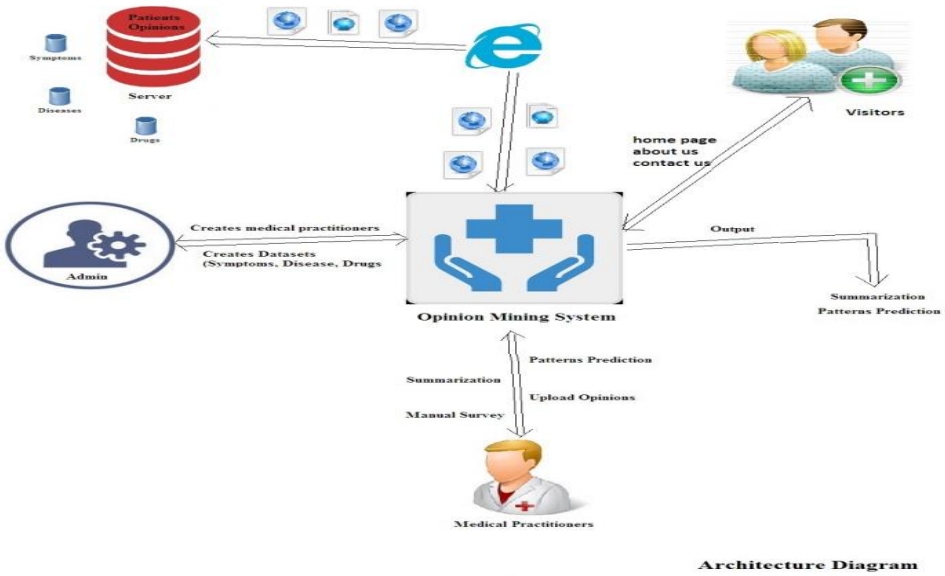


Figure 1: System Architecture

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9739

V. IMPLEMENTATION

The patients opinions are collected from reputes websites and downloaded stored as data set in opinions based system. Summarization is performed on each patients opinions and Medical pattern discovers five different patterns and identifies relationship between symptoms,disease and drugs.sentimental analysis shows satisfaction of drugs and positive and negative opinions of drugs posted by patients.visitors views patients opinions and side effects of drugs drugs posted by patients.

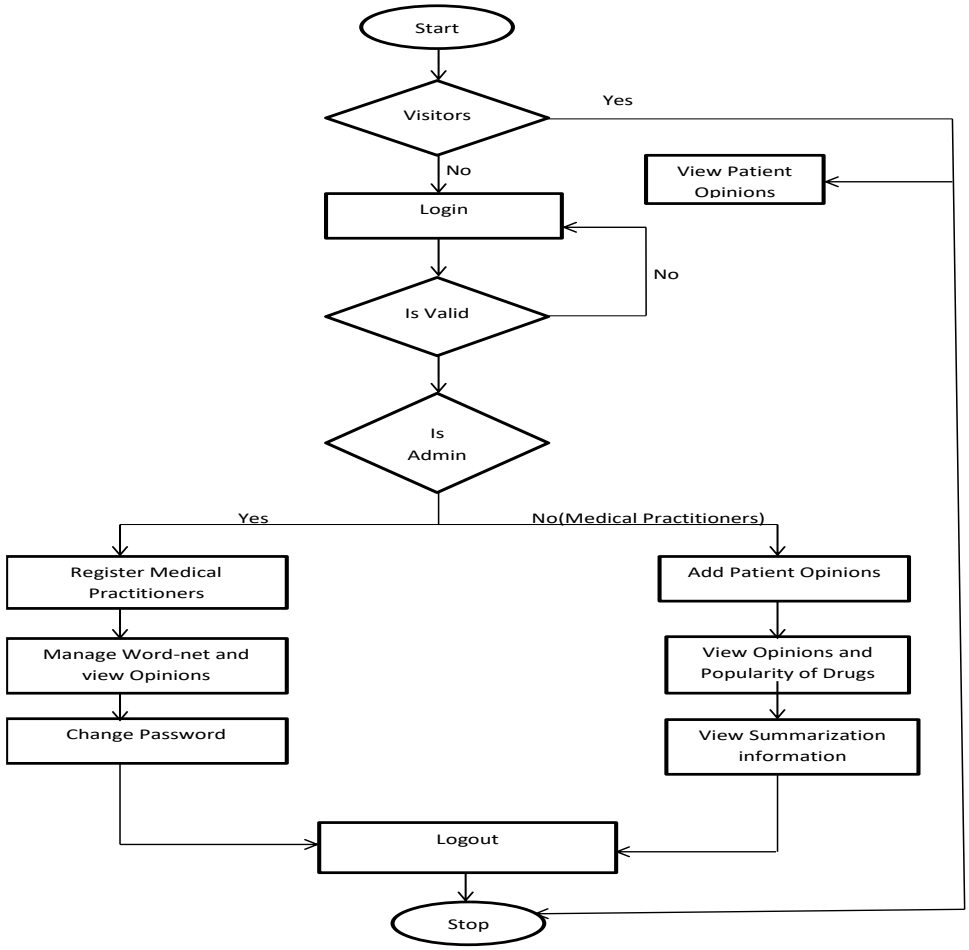


Figure 2: Workflow diagram

A. LESK BASE ALGORITHM

In Summararization is defined as Scan the Opinion Database (retrieval of all Patient Opinions)and Scan word-net (collection of all symptoms,diseases,drugs) and extract keywords from patients opinions.

Algorithm Steps

STEP 1:For each entry UI[Opinions]in buffer[storage server] do

STEP 2:Trace all keywords using the following steps

STEP 3:Tokenization [keyword extraction method-removing the stop words and retrieving the keywords] Remove punctuation,special characters,numbers etc…

STEP 4: Clustering the keywords(grouping of similar objects) By comparing with the predefined data-set(created by admin) String comparison and identity the symptoms,diseases,drugs and positive and negative words.

STEP 5: Output -Summarization Results.

B. APRIORI ALGORITHM

In the project we use " Apriori Algorithm" to find the relationship between symptoms, diseases and drugs. Eclat algorithm is one of the efficient algorithm and takes less time for data processing. This algorithm works fine for small data-sets as well as large data-sets.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9739

Pattern Prediction Process

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: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.We use Apriori algorithm to process data and to find the patterns.

Apriori 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.

Apriori Algorithm Steps:

STEP 1: Scan the data set and determine the support(s) of each item.

STEP 2: Generate L1 (Frequent one item set).

STEP 3: Use Lk-1, join Lk-1 to generate the set of candidate k - item set.

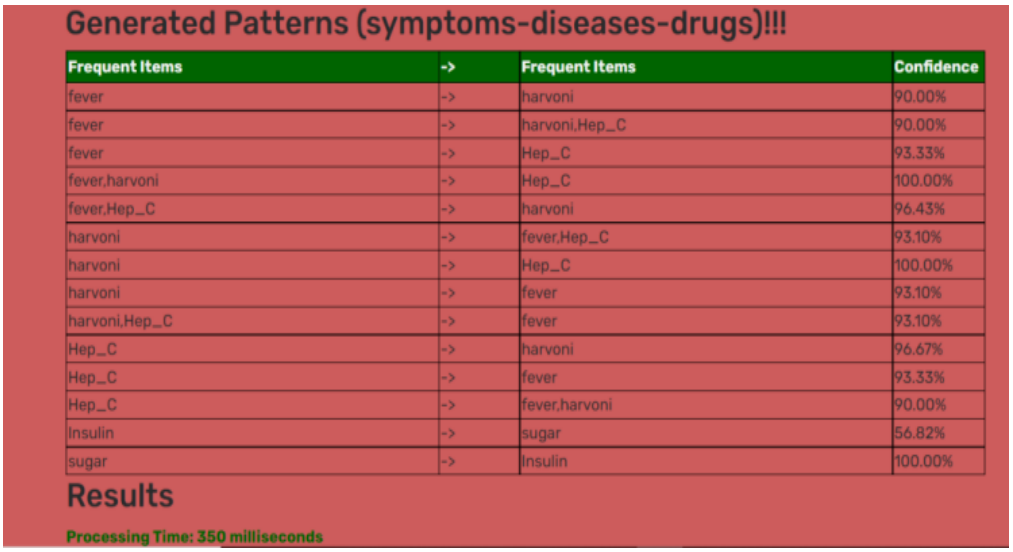
STEP 4: Scan the candidate k item set and generate the support of each candidate k – item set.

STEP 5: Add to frequent item set, until C=Null Set.

STEP 6: For each item in the frequent item set generate all non empty subsets.

STEP 7: For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence .Then add to Strong Association Rule.

Apriori Results :



Generated Patterns (symptoms-diseases-drugs)!!!

Frequent Items	->	Frequent Items	Confidence
fever	->	harvoni	90.00%
fever	->	harvoni,Hep_C	90.00%
fever	->	Hep_C	93.33%
fever,harvoni	->	Hep_C	100.00%
fever,Hep_C	->	harvoni	96.43%
harvoni	->	fever,Hep_C	93.10%
harvoni	->	Hep_C	100.00%
harvoni	->	fever	93.10%
harvoni,Hep_C	->	fever	93.10%
Hep_C	->	harvoni	96.67%
Hep_C	->	fever	93.33%
Hep_C	->	fever,harvoni	90.00%
Insulin	->	sugar	56.82%
sugar	->	Insulin	100.00%

Results

Processing Time: 350 milliseconds

CONCLUSION

In this work,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

DOI: 10.17148/IARJSET.2022.9739

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.

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.