IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207064|

A Smart Medical System to Identify the Side Effects of Drugs

Neha P, Ms. Bhavani R

Student, Department of MCA, Visvesvaraya Technological University, The National Institute of Engineering, Mysore, India

Assistant Professor, Department of MCA, Visvesvaraya Technological University, The National Institute of Engineering, Mysore, 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.

KEYWORDS: Data Science, Symptoms, Diseases, Drugs, Machine Learning, Eclat, Lesk Based, Algorithm

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, 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 is likely to occur whenever the LHS set occurs [3].

II. LITERATURE SURVEY

IEEE PAPER TITLE: Association Rule Mining on Type 2 Diabetes using FP-growth association rule
YEAR OF PUBLICATION: 2013
AUTHORS: Nandita Rane1, Madhuri Rao
METHODOLOGY: FP growth algorithm used
LIMITATIONS: this paper mainly concentrates on only diabetes disease and related patterns.
IEEE PAPER TITLE: A Comparative approach for Sentiment Analysis from Summarized User Health posts
YEAR OF PUBLICATION: 2016
AUTHORS: Mr Ajay A V, Dr. Chethan H K
METHODOLOGY: NLP techniques used
LIMITATIONS: this paper aims at comparative analysis of different sentiment analysis techniques which is used to process the patients' opinions. Doesn't predict the relationship b/w symptoms-diseases-drugs.
IEEE PAPER TITLE: Knowledge discovery from user health posts
YEAR OF PUBLICATION: 2015
AUTHORS: Vinod L. Mane, Suja S. Panicker, Vidya B. Patil
Methodology USED: AIT algorithm, FP growth algorithms and apriori used

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207064|

LIMITATIONS: the used algorithms take more time for processing, less efficient.

III. 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 performs 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 performs knowledge discovery from user posts, whereby useful `patterns about the triad `drugs-symptoms-medicine' is done by Association Rule Mining.

IV. METHODOLOGY

4.1 Unsupervised Learning

Step 1: In the 1st 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 pre-processed 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 pre-processed patients' opinions are summarized using lesk based algorithm and output of lesk based algorithm is inputted to Éclat algorithm to discover the medical patterns (symptoms-diseases-drugs).
Step 4: Éclat 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 analysed and represented visually

V. PSEUDO CODE

Eclat Algorithm Pseudo-code

```
double.Parse(lv_Transactions.Items.Count.ToString()) / 20;
        double _supportCnt = (double.Parse(lv_Transactions.Items.Count.ToString()) / 100) * 10;
        double _confidence = 0.9;
        //function to calculate the L1 and store in buffer
        Dictionary<string, string> _FrequentItemsMain = L1(_supportCnt);
        Dictionary<string, string> _Candidates = new Dictionary<string, string>();
        do
        {
           //function to calculate C2,C3.... and store in buffer
           _Candidates = GenerateCandidates(_FrequentItemsMain);
           //function to calculate L2,L3.... and store in buffer
           _FrequentItemsMain = GetFrequentItems(_Candidates, _supportCnt);
        }
        while (_Candidates.Count != 0);
        //function to generate rules or patterns
        List<ClassRules> RulesList = GenerateRules();
        //compare with the confidence and find strong rules
        List<ClassRules> StrongRules = GetStrongRules(_confidence, RulesList);
        //function to display the final results (L and Patterns)
        Result(DictionaryAllFrequentItems, StrongRules);
```

VI. RESULTS

ECLAT ALGORITHM
Performance Factor
Data Structure – array based
Memory Utilization – depends on the data set [less for small datasets]
No.of. scans - single scan required
Execution time - execution time depends on producing candidates

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1207064|

No of Instances (records)	Execution Time (milli Secs)
Around 2k	1557
Around 1k	1295
Around 500 records	865
100 records	445

Summarization Module (View)
This is the core module where system summarize all user posts and come out with useful information or conclusions. Here we make use of keyword extraction method to extract symptoms-diseases-drugs from the user health posts.
Pattern Prediction Module (View)
Here system discovers useful patterns based on side-effects per drug posted by the users of the application. This the core module of the project which finds the correlation between symptoms- diseases-drugs. To predict relationship, we use "*Eclat Algorithm*" which is an efficient algorithm to discover patterns.
SA Module
In this module patients gives their rating on a particular drug. Based on this we can track the top most drug and details.

VII. CONCLUSION AND FUTURE WORK

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.

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 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 ofSystems Engineering (SOSE), Adelaide, Australia- June 9-13, pp 85-90, 2014.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details