International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Application of Association Rule Mining in Identifying Sequence of Viral Infective Diseases

Prof. Deven Mahajan Indira College of Commerce and Science Pune, India [1]
Prof. Shital Pashankar Indira College of Commerce and Science Pune, India [2]

Abstract: Data mining is used to find out the meaningful information from the large dataset. Data mining plays vital role in various domains including healthcare, banking, retail, education, telecommunication, information security, finance etc. Medical Health Records serves as rich knowledge sources for data mining. The availability of enormous amounts of medical data can be utilized to extract constructive facts using various data mining techniques. Numerous researches have been conducted in the field of medical data mining. In this paper, a novel approach is suggested to extract association rules from medical records by choosing the best association rule mining algorithm using multiple-criteria decision analysis. The purpose of the project is to discover correlation between diseases, diseases and symptoms, diseases

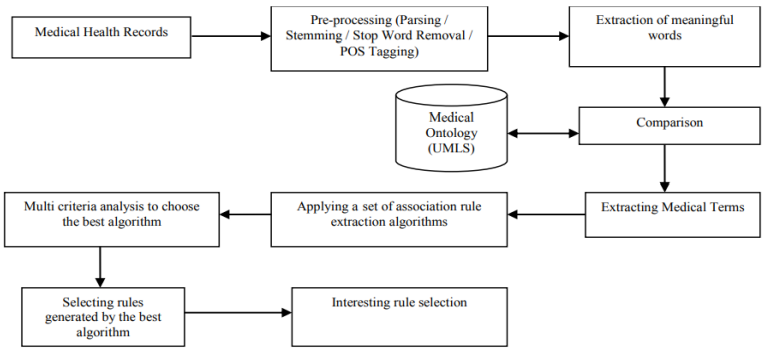
Keywords: Data Mining, Association Rule, Apriori Algorithm, Multi-criteria decision analysis, Diseases.

I. INTRODUCTION

In data mining, Association Rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness. Before applying Association rule data needs to be pre processed

System Architecture The overall architecture of the proposed system is shown in Fig.1. The dataset consists of 309 medical records which are medical transcription files. Natural Language Processing techniques are used for the extraction of medical terms

Diagram 1 System Architecture



The overall architecture of the proposed system is shown in Diagram .1. The dataset consists of medical records which are medical transcription files. Natural Language Processing techniques are used for the extraction of medical terms.

Based on the concept of strong rules, we can introduce association rules for discovering the symptoms of particular diseases for healthcare as {Fever, Cough, Chill, Bodyache}➔ {Influenza-Flu} this rule indicates that if the symptoms are Fever, Cough, Chill, Bodyache then a person is having Influenza-Flu disease; this kind of information is useful for healthcare, medical experts for the better decision making and proper treatment.

II. ASSOCIATION RULE MINING

Association rule is having two main important properties namely Support and Confidence as
Support (AB) = P (AUB)
Confidence (AB) = P (B|A)
If we correlate support and confidence then
Confidence (AB) = P (B|A) = Support_ count (AUB)/Support_ count (A)
Support_ count (AUB): number of transaction containing the item sets AU B,
Support_ count (A): number of transactions containing the item set A. [2]

Apriori Algorithm: (Finding Frequent Item sets Using Candidate Generation)

Apriori is an influential algorithm for mining frequent item sets. The name of the algorithms is based on the fact that the algorithm uses prior knowledge of frequent item sets properties. Apriori employs an iterative approach known as a level-wise search.

Input:
database D
Mini Support ϵ
Mini Confidence ξ
Output:
R_t All association rules
Method:
01 L_1 = large 1-itemsets;
02 for(k=2; $L_{k-1} \neq \emptyset$; k++) do begin
03 C_k =apriori-gen(L_{k-1}); //generate new candidates from L_{k-1}
04 for all transactions T $\in$ D do begin
05 C_t=subset(C_k,T); //candidates contained in T.
06 for all candidates C $\in C_t$ do
07 Count(C)=Count(C)+1; // increase support count of C by 1
08 end
09 L_k={C $\in C_t$ | Count(C) $\geq \epsilon \times$ | D | }
10 end
11 $L_f = \bigcup_k L_k$;
12 R_t=GenerateRules(L_f, ξ)

Diagram 2 : Apriori Algorithm

[4] III.LITERATURE REVIEW

The researcher Doddi S, et.al had worked on association rule generation of medical data containing the patient's records and reported diagnosis. An Apriori algorithm is used for association rule generation of medical data. [1]
The research paper entitled "Association Rules Mining

The author Jagmeet Kaur and Neena Madan entitled a research paper "Association Rule Mining: A Survey", explained about the association rule theoretical views along with an Apriori algorithm, FP growth algorithm and genetic algorithm. The authors also discussed about the application of association rule such as Market Basket Analysis, CRM of the Credit Card Business, Medical Diagnosis, Census Data, Protein Sequences, etc. [5]

A research paper entitled "ASSOCIATION RULE MINING ON MEDICAL DATA TO PREDICT HEART DISEASE" by Said I., et.al illustrated association rule generation for medical data for prediction of heart diseases using Apriori algorithm. The researcher had used medical data through which the sick and healthy patient's details are categorized for the prediction of heart diseases. [7]

IV. ASSOCIATION RULE GENERATION FOR VIRAL INFECTIVE DISEASES

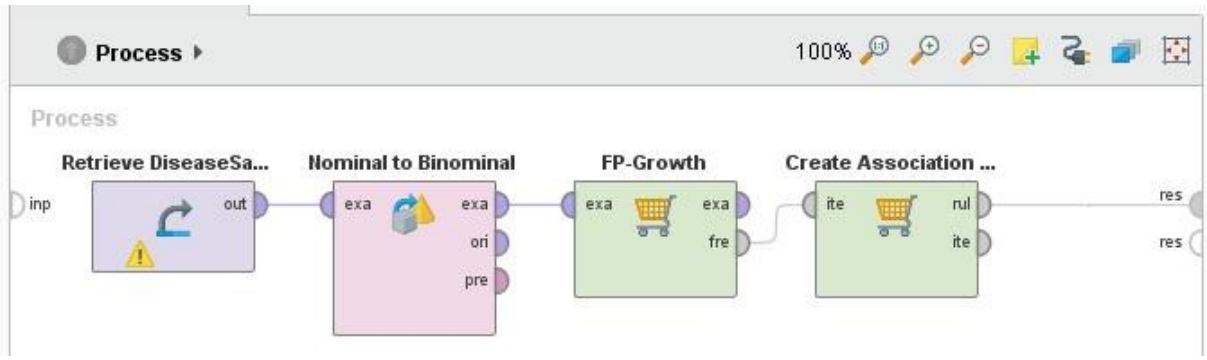


Diagram 3: Dataset for Association Mining in Rapid Miner

This research used an Apriori algorithm and FP-Growth an association rule generation algorithm for various viral infective diseases and their symptoms. The analysis of the diseases dataset is done using Rapid miner text mining tool. The Diagram.2 represents an association rule generation for viral infective diseases. The data source is retrieved through excel which contains diseases and their symptoms then he data is converted from nominal to binomial. The FP- Growth technique is used to set the min support and confidence and then association rules are generated. The association rules

generated through the Rapid miner are represented in Diagram 3.
For example first rule indicates that

{swelling, fever, headache, feeling-sick, dry-mouth, joint-ache}➔ {Mumps}
{Mumps} ➔ {swelling, fever, headache, feeling-sick, dry-mouth, joint-ache}

It will be helpful for various hidden pattern generation and interpretation for further decision making and making healthcare sector helpful.

Result in Rapid Miner:

Association Rules
[Symptoms = swelling, fever, headache, feeling-sick, dry-mouth, joint-ache] --> [Disease = Mumps] (confidence: 1.000)
[Disease = Mumps] --> [Symptoms = swelling, fever, headache, feeling-sick, dry-mouth, joint-ache] (confidence: 1.000)
[Symptoms = fever, chills, cough, soar-throat, runny-nose, bodyache, headache, fatigue, vomitting] --> [Disease = InfluenzaFlu] (confidence: 1.000)
[Disease = InfluenzaFlu] --> [Symptoms = fever, chills, cough, soar-throat, runny-nose, bodyache, headache, fatigue, vomitting] (confidence: 1.000)
[Symptoms = fever ,jointpain, headache, musclepain, joint-swelling, rash] --> [Disease = Chikungunya] (confidence: 1.000)
[Disease = Chikungunya] --> [Symptoms = fever ,jointpain, headache, musclepain, joint-swelling, rash] (confidence: 1.000)
[Symptoms = cough,coryza,runny-nose, fever, sneezing, rash, boadyache, watery-eyes, Soar-Throat] --> [Disease = Measles] (confidence: 1.000)
[Disease = Measles] --> [Symptoms = cough,coryza,runny-nose, fever, sneezing, rash, boadyache, watery-eyes, Soar-Throat] (confidence: 1.000)
[Symptoms = chills,fever,coughing,sore-throat,runny-stuffy-nose,bodyaches,fatigue, nausea,vomiting] --> [Disease = Swine Flu] (confidence: 1.000)
[Disease = Swine Flu] --> [Symptoms = chills,fever,coughing,sore-throat,runny-stuffy-nose,bodyaches,fatigue, nausea,vomiting] (confidence: 1.000)
[Symptoms = abdominal-pain, nausea, vomiting] --> [Disease = Diarrhoea] (confidence: 1.000)
[Disease = Diarrhoea] --> [Symptoms = abdominal-pain, nausea, vomiting] (confidence: 1.000)
[Symptoms = Rash, Red-spot, Soar-throat, high-fever, vomitting, headache, tiredness, Painfull-Blisters] --> [Disease = Chickenpox] (confidence: 1.000)
[Disease = Chickenpox] --> [Symptoms = Rash, Red-spot, Soar-throat, high-fever, vomitting, headache, tiredness, Painfull-Blisters] (confidence: 1.000)
[Symptoms = Nausea, vomiting, Dehydration, Muscle-cramps, abdominal-pain] --> [Disease = Cholera] (confidence: 1.000)
[Disease = Cholera] --> [Symptoms = Nausea, vomiting, Dehydration, Muscle-cramps, abdominal-pain] (confidence: 1.000)
[Symptoms = High-Fever, Abdomen-Pain, vomiting, cold, headache, rash] --> [Disease = Dengue] (confidence: 1.000)
[Disease = Dengue] --> [Symptoms = High-Fever, Abdomen-Pain, vomiting, cold, headache, rash] (confidence: 1.000)
[Symptoms = Fever, Headache, Nausea, Vomiting, Anxiety, Hyperactivity, Confusion, Agitation] --> [Disease = Rabies] (confidence: 1.000)
[Disease = Rabies] --> [Symptoms = Fever, Headache, Nausea, Vomiting, Anxiety, Hyperactivity, Confusion, Agitation] (confidence: 1.000)

V. CONCLUSION

The above results are drawn for the diseases information with the disease name and symptoms of diseases which are in structured form. According to the report of Oracle Data Mining Concepts Now a day's, approximate 90% of data is in the unstructured format that is text, images, audio, video, graph, email, blog, etc and only 10 % of data is in the structured form. There is need to focus on unstructured data and getting the better results from it. The future scope of the present study is to work on text mining of healthcare data and generating the various patterns from it. This research used an Apriori algorithm an association rule generation algorithm for various viral infective diseases and their symptoms

REFERENCES

[1]. Doddi S. , Marathe A. S. S. Ravi, David C. Torney, Discovery of Association Rules in Medical Data
[2]. M. Inbava lli(2015).Efficient Analysis of Frequent item set Association Rule Mining Methods. International Journal of Scientific & Engineering Research, 6(4), 1-23, http://www.ijser.org/researchpaper%5CEfficient-Analysis-of-Frequent-itemset-Association-Rule-Mining-Methods.pdf
[3]. Kotsiantis S., Kanellopoulos D.(2006).Association Rules Mining: A Recent Overview, GESTS International Transactions on Computer Science and Engineering, 32 (1), 71-82
[4]. Zhao Q. ,Bhowmick S. (2003).Association Rule Mining: A Survey. Technical Report, CAIS, Nanyang Technological University, Singapore, No. 2003116 ,1-20
[5]. Kaur J. , Madan N. (2015). Association Rule Mining: A Survey. International Journal of Hybrid Information Technology ,8(7),pp.239-242 http://dx.doi.org/10.14257/ijhit.2015.8.7.22
[6]. Rashid M. , Hoque T., Sattar A. Association Rules Mining Based Clinical Observations. Griffith University Nathan, QLD, Australia
[7]. Said I. , Haruna A. , Garko A. (2015). ASSOCIATION RULE MINING ON MEDICAL DATA TO PREDICT HEART DISEASE International