Heart Disease Prediction System using Enhanced Apriori

P. Parameswari, C.Ramachandran, R. Rassika

Abstract— Heart disease is frightening the people around the world and in some countries it is the number one disease which leads to death. Biomedical research efforts help to prevent and treat heart disease in a better way. Handling large amount of data is often very tedious with traditional methods which lead into problems, particularly in high level of complexity and vagueness factors. Mining frequent patterns from large databases has emerged as an important area in data mining research and knowledge discovery community; this also contributes so much to health care domain. This heart prediction system helps to predict heart related problems at an early stage. The proposed system predicts heart related issues of a person based on questions and the answers given to the prediction system. To have better results in minimum time duration an Enhanced Apriori algorithm was introduced which is an improvement of Apriori algorithm. The experimental results proved that the proposed approach performs faster and memory efficient with more number of patterns. It was also proved that the prediction rate of Enhanced Apriori was also good (94%) than Apriori (87%).

Keywords— Prediction, Data Mining , Heart Disease, Apriori, Association Rule Mining

I. INTRODUCTION

Data Mining is usually performed on discrepant data as there are several data mining functionalities available to find the specific patterns. The DM functionalities consist of associations, frequent pattern mining, clustering, classification and regression. Association rule mining algorithm helps to find the relationship between the large amount of data. Association rule mining algorithms are used by researchers in various domains. It helps in discovering frequent item sets that exist within a dataset. Data mining can be used in marketing to identify the items that is purchased during a single transaction. Association rules help in predictions, advertisements and health care to predict diseases at an earlier state.

II. RELATED WORKS

Fuzzy logic combined with artificial neural network was applied for knowledge engineering techniques was proposed [3]. Modified co-occurrence and cluster based mean mode method was proposed to handle mixed data types which can be used with any type of system [2]. Modified fuzzy ant-miner technique was proposed for medical diagnosis to provide accurate results [5].

Revised Version Manuscript Received on 25 November, 2018.

Dr.P.Parameswari, Department of MCA, Kumaraguru College of Technology, Coimbatore, India,

Mr.C.Ramachandran, Department of MCA, Kumaraguru College of Technology, Coimbatore, India

Ms.R. Rassika, Department of MCA, Kumaraguru College of Technology Coimbatore, India

There are number of methods used to generate rules from database in medical domain [6]. An artificial bee colony algorithm was modified to perform heart disease prediction process and modifications to this algorithm have been used with evolutionary algorithm to create classifier system that helps the physicians in decision making [7]. An approach that generates association rules on numeric data is proposed to apply in medical data. Standard Apriori and predictive Apriori algorithms finds the frequent item sets which occurs in a dataset by generating rules [8]. Apriori can work on databases that contain more number of transactions.

III. PROPOSED METHOD

Nowadays data mining contributes more for healthcare industry. It helps to use medical data for analysis and offer health care benefits at a minimum cost. Mining algorithms helps to find solutions that arise frequently and provide valuable suggestions for improvements. The data mining techniques play a major role in diagnosing of so many health issues and to predict various diseases like heart disease, diabetes, fertility ,cancer and skin disease Methodology to predict heart diseases at an early stage has been discussed in this section.

Table 1: Features used for Analysis

Attributes	Description	Type
Age	Age (years)	Nominal
Gender	Male/Female	Nominal
CP	Type of chest pain	Numeric
BP	Blood Pressure	Nominal
Chol	Serum cholesterol	Nominal
Sugar	Blood sugar	Nominal
ECG	Electro Cardio graphic Results	Numeric
Thalach	Maximum heart rate	Numeric
Disease Rate	Diagnosis of heart disease (Prediction Rate)	Nominal

A. Dataset

Categorical data related to our research was collected and binned into small categories. The cut-off values of each bins are given below,

- Age: 20–39(low), 40–59(medium), 60 plus (high)
- Gender: male, female
- CP: 1(typical angina), 2(atypical angina) 3(Non-anginal pain), 5(Asymptomatic)
- BP: 70 - 90 (low), 90 - 120 (ideal), 120 - 140 (pre high) ,140 – 190(very high)
- Chol: (> 200mg/dl – ideal), (200-239mg/dl-borderline high), (<240 mg/dl – high)
- Sugar: (70-110-normal) (>140-high)
- ECG: (Value 0: normal) (Value 1: having ST-T wave abnormality1)
- Thalach: Maximum heart rate achieved
- Disease Rate: (angiographic disease status) (low:< 50% diameter narrowing) (high: > 50% diameter narrowing)

Based on the slap values the data was converted into categorical data. The categorical data was used to predict heart disease by answering the questions related to the symptoms listed in the questionnaire and final results was produced.

B. Apriori

Apriori works based on the knowledge acquired previously .It finds frequent item sets and its properties by an iterative approach .It was also known as level-wise search. It helps in finding frequent patterns from relational databases and other heterogeneous databases.

Rule mining process:

(i) Item sets occurs recurrently was identified .

(ii) Strong association rules are generated .

Apriori algorithm:

C_ik: Candidate item sets

k: Size of item set, I = Item;

Fk: Frequent item sets ,

```
Fk =  FI
  for (k = 1; Fk != 0; k++)
  do begin
     Ck+1 = Fk;
     increment  count  in Ck+1 for each
     Transaction TR;
Fk+1 = Ck+1
 end
return Fk;
```

If minimum support count is met by the item set it can be considered as frequent item set and it is denoted by Fk.

C. Enhanced Apriori

Enhanced Apriori algorithm reduces the time for candidates' item set generation.

Algorithm:

```
F1 = frequent itemsets1 (T);

For (k = 2; Fk! =0   ; k++)

 {

Ck = Fk-1;

N=  i_minsup (Ck, F1);

TID = get_Tran_ID (x);

for each  one of the transaction T  in Tgt Do
 count  in Ck  = Ck+1 if found  ;

Fk =  Ck  with minimum support;

End;

 }
```

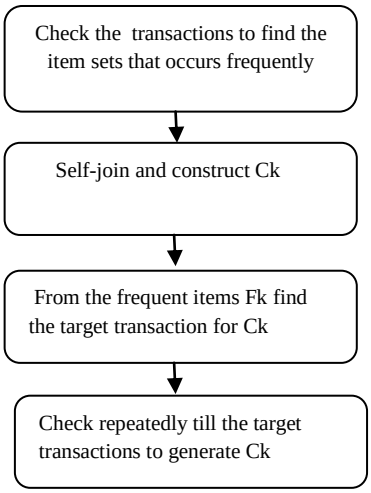

Figure :1 Enhanced Apriori Flow Diagram

D. Association Rules

Rules used to find patterns from a large database which now applied in a medical database for data analysis and this rules satisfies both minimum support and confidence levels and form strong rules.

Sample Rules Generated:

- If Age = Medium AND Sugar=High AND CP=4 AND BP=high, THEN Affected rate= High
- If Age = High AND Sex = Male AND CP = 2 AND THEN Affected rate= Low
- If Age = Medium AND Sex = Female AND Chol = very high AND BP = High AND Sugar=High, THEN Affected rate = High

IV. RESULTS AND DISCUSSION

We present the results of classical Apriori algorithm by comparing it with Enhanced Apriori algorithm. We have considered heart disease dataset for discussion and this dataset with 9 attribute values. A tool was constructed and data was collected from the users for analysis. The user should answer, the questions given in the template and the system will predict whether the person have heart disease or not. We have collected 14 inputs from the user for prediction and selectively 9 attributes are considered for analysis

Dataset: Heart disease

Attributes: 9

Number of Class Attributes: 1

Enhanced Apriori algorithm compared with classic Apriori in terms of execution time which provides good results while it was experimented with more amounts of instances.

Table 2: Performance of Apriori Algorithms in terms of Execution Time according to Minimum support count

Min_support	Apriori (S)	Enhanced Apriori (S)	Time Difference (%)
0.02	7.02	1.614	79.96%
0.04	1.99	0.413	78.57%
0.06	1.23	0.326	74.91%
0.08	0.42	0.190	53.96%
0.10	0.33	0.128	55.07%

The pictorial representation of execution time of these two algorithms is given in figure 2. X-axis represents minimum support count and Y-axis represents the execution time. Where we can see the time difference between both the algorithms is high and Enhanced Apriori performs better than the classical Apriori.

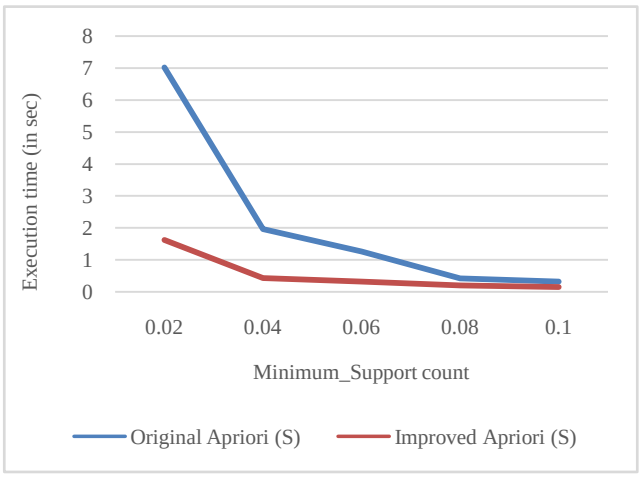

Figure 2: Performance of Apriori Algorithms in terms of Execution Time

V. CONCLUSIONS

Regarding the high occurrence of heart diseases and its consequence of mortality, some intelligence systems should be developed to manage it. So the proposed system can help people to predict heart diseases at an early stage by their symptoms. This research comes out with prediction of disease by generating rules by using Enhanced Apriori (EA) algorithm with minimum execution time and with better results comparing to the traditional algorithm .Pre-processing helps to remove unwanted data and retain the data quality. The proposed system shows that the time complexity of Enhanced Apriori (EA) algorithm is low comparing with classic Apriori because of minimum number of iterations. When classic Apriori is used for diagnosing diabetics the prediction rate was 87 % and for Enhanced Apriori the prediction rate is 94 %.

REFERENCES

1. Wilson, F. N. 1948. The clinical value of chest leads. British Heart Journal 10 (2): 88–91.
2. Parameswari, P, Abdul Samath, J & Saranya, S 2015, Efficient birch clustering algorithm for categorical and numerical data using modified co-occurrence method, International Journal of Applied Engineering
3. Research, vol. 10, no. 11, pp. 27661-27673 .
4. Truett J. J, Cornfield, and W. B. Kannel. 1967. A multivariate analysis of the risk of coronary heart disease in Framingham. Journal of Chronic Diseases 20:511-24.
5. Rose. G. and H. Blackburn. 1968. Cardiovascular survey methods: WHO technical report series no. 56. Geneva: World Health Organization..
6. Aribarg, T., Supratid, S., Lursinsap, C, 2012. Optimizing the modified fuzzy ant-miner for efficient medical diagnosis. Applied Intelligence ;37(3):357–376.
7. Silverstein, C., Brin, S., Motwani, R. and Ullman, J, 1998. Scalable Techniques for Mining Causal Structures, Technical Report, Department of Computer Science, Stanford University.
8. Luepker, R. V., A. Evans, P. McKeigue, and K. S. Reddy, 2004. Cardiovascular Survey Methods. 3rd edition Geneva: World Health Organization.
9. Blair S. N., J. B. Kampert, H. W. Kohl, C. E. Barlow, C. A. Macera, R. S. Paffenbarger, and L. W. Gibbons, 1996. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. Journal of the American Medical Association 276: 205-210.
10. Rakesh Agrawal and Ramakrishnan Srikant , 1994. Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September.