© 2016 IJSRST | Volume 2 | Issue 3 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Finding Association Rules in Medical Datasets

Jasmeen Kaur Chahal
Department of Computer Science & Information Technology, Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, Punjab, India

ABSTRACT

Association Rule mining is one of the most important fields in data mining and knowledge discovery. In this paper we define an algorithm which associates the symptoms of the patient and defines the disease of the patient. By taking the values of minimum support and confidence the algorithm gives the result.
Keywords: Association Rule Mining, Support, Confidence.

I. INTRODUCTION

The field of data mining has been growing rapidly due to its broad applicability, achievements and scientific progress. It is the process of knowledge discovery where knowledge is gained by analyzing the data store in very large repositories, which are analyzed from various perspectives and the result is summarized it into useful information [1].The overall aim of data mining process is to extract information from a dataset and transform it into an understandable structure for further use. There are number of fields where data mining techniques are used for efficient discovery of valuable, non-obvious information from large collection of data. One of the main techniques of data mining is Association Rule Mining (ARM). It is the popular technology and well used by various researchers to finding frequent patterns in large databases. Association rules are created by analyzing data for frequent if/then patterns and using the criteria *support* and *confidence* to identify the most important relationships. *Support* is an indication of how frequently the items appear in the database. *Confidence* indicates the number of times the if/then statements have been found to be true.

There are number of fields where implementation of association rule mining has been done to find useful data for effective decision making: **Market basket analysis**, Managers could use this data for adjusting store layouts (placing items optimally with respect to each other), for cross-selling, for promotions, for catalog design and to identify customer segments based on buying patterns**.** **Medical diagnosis**, applying association rules in medical diagnosis can be used for assisting physicians to cure patients. **CRM of credit card business**, Customer Relationship Management (CRM), through which, banks hope to identify the preference of different customer groups, products and services tailored to their liking to enhance the cohesion between credit card customers and the bank, has become a topic of great interest [2]. **Census data,** Censuses make a huge variety of general statistical information on society available to both researchers and the general public [3]. The information related to population and economic census can be forecasted in planning public services(education, health, transport, funds) as well as in public business(for setup new factories, shopping malls or banks and even marketing particular products). As mining association rule in medical field is helpful for physicians to diagnose the patients. The main goal of this work is to find the patients, which has frequent similar symptoms from database using proposed algorithm and identify a disease. Section II: explores the different algorithms to mine association rules. Section III: the proposed algorithm used in this work. Section IV presents results. Section V states the conclusion.

II. METHODS AND MATERIAL

A. Algorithms of Association Rule Mining:

AIS Algorithm: The AIS algorithm [4] was the first algorithm proposed by Agrawal, Imielinski, and Swami for mining association rule. It focuses on improving the quality of databases together with necessary

functionality to process decision support queries. In this algorithm only one item consequent association rules are generated, which means that the consequent of those rules only contain one item, for example, rules like $X \cap Y \Rightarrow Z$ can be generated but not the rules like $X \Rightarrow Y \cap Z$.

SETM Algorithm: In the SETM algorithm, candidate itemsets [5] are generated on-the-fly as the database is scanned, but counted at the end of the pass. Then new candidate itemsets are generated the same way as in AIS algorithm, but the transaction identifier TID of the generating transaction is saved with the candidate itemset in a sequential structure.

APRIORI Algorithm: In this algorithm, frequent subsets are extended one item at a time and this step is known as candidate generation process. Then groups of candidates are tested against the data. To count candidate item sets efficiently, Apriori uses breadth-first search method and a hash tree structure. It is submitted by Agarwal and R.Srikant[6] in 1994 is the most effective algorithm of mining association rules.

APRIORITID Algorithm: In this algorithm [7], database is not used for counting the support of candidate itemsets after the first pass. The process of candidate itemset generation is same like the Apriori algorithm.

APRIORIHYBRID Algorithm: As Apriori does better than Aprioritid in the earlier passes and Aprioritid does better than Apriori in the later passes. A new algorithm [7] is designed that is Apriorihybrid which uses features of both the above algorithms. It uses Apriori algorithm in earlier passes and Aprioritid algorithm in later passes.

FP-GROWTH Algorithm : FP-growth requires constructing FP-tree[8]. For that, it requires two passes. FPgrowth uses divide and conquer strategy. It requires two scans on the database. It first computes a list of frequent items sorted by frequency in descending order (F-List) and during its first database scan. In the second scan, the database is compressed into a FP-tree [9].

B. Proposed Algorithm

As discussed above algorithms, the Apriori algorithm is most effective algorithm of mining association rules. Its original motivation is to analyze the market basket and it aims to find out the association rules between different goods in the transaction database. This algorithm has wide range of applications. It is used by Yan Peng et al.[10] used algorithm in extracting the key factors in many elements that affect the enterprises' strategic decision making.

Also J.Manimaran et al.[11] discuss that Apriori algorithm is suitable algorithm to analyze unknown information available in text data. Santhana Joyce m et al.[12] uses Apriori algorithm as base algorithm and defines new algorithm FDM to find association rules in horizontally distributed databases.

In this work the (EAA) Enhanced Apriori Algorithm has been used. EAA is based upon two values i.e Support and Confidence Values. The steps are:

1) Scan the transaction database to get the support of S each 1-itemset, compare S with min_sup, and get a support of 1-itemsets,L1.
2) Use L_{k-1} join L_{k-1} to generate a set of candidate k-itemsets. And use Apriori Property to prune the unfrequented k-itemsets from this set.
3) Scan the transaction database to get the support S of each candidate k-itemsets in the find set, compare S with min_sup, and get a set of frequent k-itemsets $L_{k.}$
4) Find that the candidate set is null or not: a) if yes, then for each frequent itemsets 1, generate all nonempty subsets of 1. b) if not, repeat the procedure from step 2.
5) For every nonempty subset s of 1, output the rule "s=>(1-s)" if confidence C of the rule "s=>(1-s)" (=support s of 1/support S of s)' min_conf.

III. RESULT AND DISCUSSION

Input Parameters:

Id No.	Attribute
1	Age
2	alchohol_intake
3	blood_cholestrol
4	physical_activity

5	Hereditary
6	Smoking
7	Diabetes
8	Diet

We apply the association rule mining over the heart database. The database taken from uci repository and the requirement of the system type for the implementation of desired algorithm in MatLab version 2015. The result obtain is shown below:

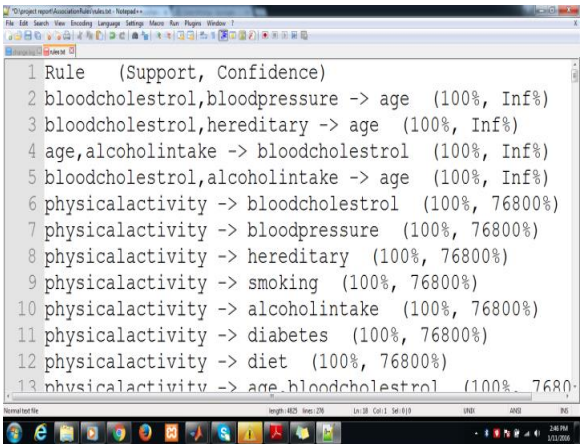

Fig. (1)

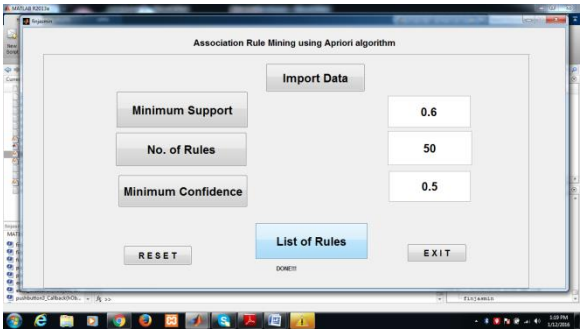

Fig. (2)

From the analysis of the result we can say that a person's heart's condition is mostly depends on the blood cholestrol & blood pressure, as well as age and alchohol_intake. While the smoking activities & physical activities doesn't effects the health conditions that much as compared to the other attributes.

IV. CONCLUSION

The major contribution of this paper is to find the association rules from the database of heart disease by using Enhanced Apriori algorithm (EAA).

V. REFERENCES

[1] Dharminder Kumar and Deepak Bhardwaj, "Rise of Data Mining: Current and Future Application Areas ", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011.

[2] R. S. Chen, R. C. Wu and J. Y. Chen, "Data Mining Application in Customer Relationship Management Of Credit Card Business", In Proceedings of 29th Annual International Computer Software and Applications Conference (COMPSAC'05), Volume 2, pages 39-40.

[3] D. Malerba, F. Esposito and F.A. Lisi, "Mining spatial association rules in census data", In Proceedings of Joint Conf. on "New Techniques and Technologies for Statistcs and Exchange of Technology and Know-how", 2001.

[4] Qiankun Zhao, Sourav S. Bhowmick, Association Rule Mining: A Survey, Technical Report, CAIS, Nanyang Technological University, Singapore, 2003.

[5] Komal Khurana, Mrs. Simple Sharma, A Comparative Analysis of Association Rules Mining Algorithms, International Journal of Scientific and Research Publications, Volume 3, Issue 5 , May 2013 ISSN 2250-3153.

[6] R. Agarwal and R. Srikant, Fast algorithms for mining association rules in large databases.In Research Report RJ 9839,IBM Almaden Research Center, San Jose, CA, June 1994.

[7] Manisha Girotra, Kanika Nagpal Saloni inocha Neha Sharma Comparative Survey on Association Rule Mining Algorithms, International Journal of Computer Applications (0975 – 8887) Volume 84 – No 10, December 2013.

[8] Sotiris Kotsiantis, Dimitris Kanellopoulos, AssociationRules Mining: A Recent Overview, GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 71-82

[9] Gagandeep Kaur, Shruti Aggarwal , Performance Analysis of Association Rule Mining Algorithms, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8, August 2013, ISSN: 2277 128X.
[10] Yen Peng, Tian Zhou , Research on Apriori Algorithm in Extracting The Key Factor, Proceedings of IEEE CCIS2012.
[11] J. Manimaran, T. Velmurgan , IEEE International Conference on Computational Intelligence and Computing Research,2013.
[12] Santhana Joyce M, Nirmalrani V, IEEE International Conference on Circuit, Power and Computer Technologies, 2015.