Vol 2 Issue 6 March 2013

ISSN No : 2249-894X

Monthly Multidisciplinary Research Journal

Review Of Research Journal

Chief Editors

Welcome to Review Of Research

RNI MAHMUL/2011/38595 **ISSN No.2249-894X**

Review Of Research Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial Board readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.

Advisory Board

Address:-Ashok Yakkaldevi 258/34, Raviwar Peth, Solapur - 413 005 Maharashtra, India
Cell : 9595 359 435, Ph No: 02172372010 Email: ayisrj@yahoo.in Website: www.isrj.net

Review Of Research
Vol.2, Issue.4, Jan. 2013
ISSN:-2249-894X
Available online at www.reviewofresearch.net

ORIGINAL ARTICLE

TO INTERACTIVE AND TO PREDICT THE DIFFERENT METHODS ASSOCIATION AND DECISION MAKING PROCESS TOOLS

LALITA

Associate professor
Manav Instiitute Technology and Manangent ,JEVRA(HISAR)

Abstract:

An interactive association rule method has been proposed to improve upon the quality of the data mining results in large medical datasets. Very often association rule mining yields numerous meaningless rules. Domain expert can often guide and restrict the search procedure to contain the number of useless rules. The association rule method proposed herein is "Apriori" based that takes a target attribute from a clinician to identify the class of patients in hepatitis database. The clinical data derived from the electronic healthcare records and the biological information derived from data mining techniques could provide more imperative input to the decision making process. The aim of the experiments reported in this chapter was to demonstrate how to mine useful knowledge hidden in the form of association rules that can help the clinicians to quickly make sense out of vast clinical datasets.

KEYWORDS:

Data mining techniques are largely used in medical databases

INTRODUCTION

Identification of patient's class in health care management is of vital importance. Clinical trails provide a method for evaluating the effectiveness and safety of new treatment for different diseases on human subjects [Bethel C.L. et.al., 2006]. Valuable knowledge can be discovered from the healthcare datasets by using the data mining techniques. Data mining and interactive decision support tools can help the clinicians to process a huge amount of data available from previous cases and help in diagnostics study to suggest probable ailment based on the values of several important attributes of patient [Podgorelec V. et.al., 2005]. Very often there exist certain attributes that play vital role in the process of diagnosis. Association rule mining finds relationship between data attributes with a given support and confidence threshold. This chapter describes the interactive use of association rule mining in the prediction of the class of hepatitis patients. The basic idea is to generate implication rules to determine the attributes that derive the class of the patients. Such rules can provide the matching criteria that can find matches for new patient records.

Hepatitis is a disease in which tissues of liver are inflamed by the hepatitis virus [Leung K.S. et.al., 2009]. A viral hepatitis is a potential risk to liver cirrhosis and hepatocellular carcinoma (HCC) which is the most common type of liver cancer and the exact cause of HCC is still unknown. Study of viral hepatitis is crucial in medicine. Hepatitis impairs liver function such as removal of harmful substances, regulation of blood composition and production of bile to help digestion. A well known sign of hepatitis is jaundice, which is an accumulation of a chemical called bilirubin in the body tissues. The symptoms of hepatitis include: a) nausea, b) vomiting, c) low grade fever, d) loss of appetite, e) rash, f) fatigue, g) pain in the liver, and h) dark brownish urine. Hepatitis can be cured if treated early, but it may cause damage if the infection persists for long.

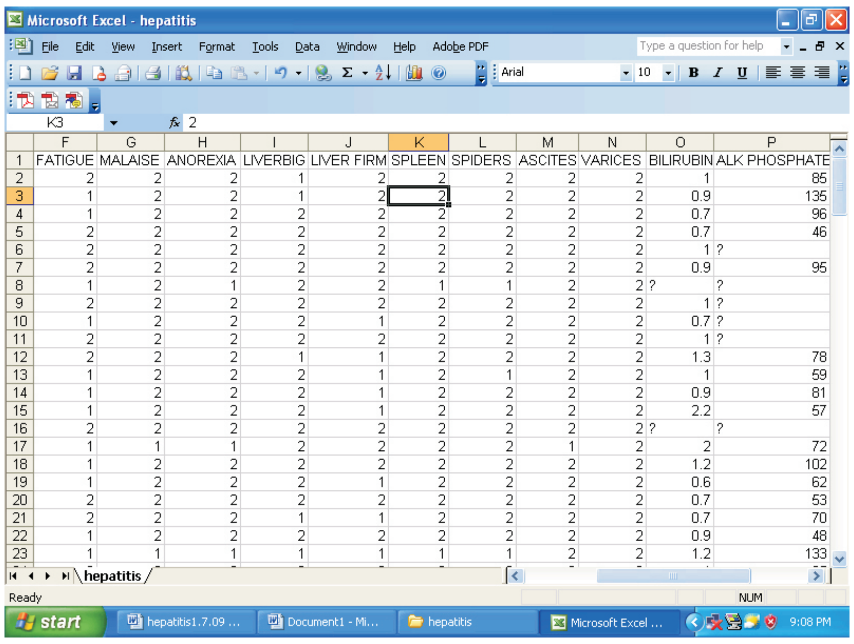
Data mining techniques are largely used in medical databases to improve the decision making in health-care. [Ho T.B. et.al., 2007] have described a temporal abstraction approach to mining knowledge from the hepatitis database of Chiba University hospital (1982-2001) to find short-term and long-term changed tests. [Leung, K.S., et.al., 2009] have advocated the application of techniques like clustering, feature selection, classification and fuzzy measure to HBV DNA sequence database of 200 Hepatitis B patients to find new and interesting knowledge for medical practitioners. [Ho T.B. et.al., 2007] has outlined a novel approach to temporal abstraction for detection and exploitation of temporal patterns and relationships between events in viral hepatitis such as "event A occurred slightly before event B and event B ended in synchrony with event C" have been presented.

An experiment has been conducted on dataset relating to hepatitis disease that contains the results of hepatitis test and some other parameters relating to the patients. Association rule mining technique has been used to predict the class of hepatitis patients. The efforts have been made to use the technique interactively so as to extract useful knowledge to improve decision making in health care. Aim of the study was to find the causes of hepatitis so that the condition of the patient might be controlled before reaching at the chronic stage.

CLASSIFICATION USING INTERACTIVE ASSOCIATION RULE MINING – A CASE STUDY OF HEPATITIS DATASET

Attribute Name	**Category**	**Possible Values**
Class	Boolean	DIE, LIVE
Age	Discrete	10, 20, 30, 40, 50, 60, 70, 80
Sex	Discrete	Male, Female
Steroid	Boolean	No, Yes
Antiviral	Boolean	No, Yes
Fatigue	Boolean	No, Yes
Malaise	Boolean	No, Yes
Anorexia	Boolean	No, Yes
Liver Big	Boolean	No, Yes
Liver Firm	Boolean	No, Yes
Spleen Palpable	Boolean	No, Yes
Spiders	Boolean	No, Yes
Ascites	Boolean	No, Yes
Varies	Boolean	No, Yes
Bilirubin	Discrete	0.39, 0.80, 1.20, 2.00, 3.00, 4.00
Alk Phosphate	Discrete	33, 80, 120, 160, 200, 250
SGOT	Discrete	13, 100, 200, 300, 400, 500
Albumin	Discrete	2.1, 3.0, 3.8, 4.5, 5.0, 6.0
Proteome	Discrete	10, 20, 30, 40, 50, 60, 70, 80, 90
Histology	Boolean	No, Yes

Table 1.1: Structure of hepatitis dataset.
(Source: Hepatitis Dataset - http://archive.ics.uci.edu:80/ml/datasets.html



	F	G	H	I	J	K	L	M	N	O	P
1	FATIGUE	MALAISE	ANOREXIA	LIVERBIG	LIVER FIRM	SPLEEN	SPIDERS	ASCITES	VARICES	BILIRUBIN	ALK PHOSPHATE
2	2	2	2	1	2	2	2	2	2	1	85
3	1	2	2	1	2	2	2	2	2	0.9	135
4	1	2	2	2	2	2	2	2	2	0.7	96
5	2	2	2	2	2	2	2	2	2	0.7	46
6	2	2	2	2	2	2	2	2	2	1	?
7	2	2	2	2	2	2	2	2	2	0.9	95
8	1	2	1	2	2	1	1	2	2	?	?
9	2	2	2	2	2	2	2	2	2	1	?
10	1	2	2	2	1	2	2	2	2	0.7	?
11	2	2	2	2	2	2	2	2	2	1	?
12	2	2	2	1	1	2	2	2	2	1.3	78
13	1	2	2	2	1	2	1	2	2	1	59
14	1	2	2	2	1	2	2	2	2	0.9	81
15	1	2	2	2	1	2	2	2	2	2.2	57
16	2	2	2	2	2	2	2	2	2	?	?
17	1	1	1	2	2	2	2	1	2	2	72
18	1	2	2	2	2	2	2	2	2	1.2	102
19	1	2	2	2	1	2	2	2	2	0.6	62
20	2	2	2	2	2	2	2	2	2	0.7	53
21	2	2	2	1	1	2	2	2	2	0.7	70
22	1	2	2	2	2	2	2	2	2	0.9	48
23	1	1	1	1	1	1	1	2	2	1.2	133

Figure 1.1: A snapshot of the hepatitis dataset converted into MS-EXCEL format.

INTERACTIVE ASSOCIATION RULE MINING

Association rule mining provides valuable information in terms of significant correlations between different attributes' values that might not be evident at the first glance in large datasets. The experimental part of this work has demonstrated benefits of integration of interactivity in Apriori approach for discovering association rules hidden in the target dataset. The interactive algorithm for discovering association rules starts by asking user's requirement with respect to attributes to be included in the search. Since the dataset has one class attribute that determines the patient class (LIVE or DIE), the clinicians are interested in finding rules that determine the value of patient class (LIVE or DIE). In addition to attribute specification, the user supplies the minimum support and confidence threshold, the two parameters required by Apriori algorithm. In the experimental runs, minimum support and confidence threshold have been fixed at 15% and 80%, respectively.

The proposed algorithm has been implemented in java environment (J2SDK1.4.1). The algorithm works in two steps: a) finds all the frequent itemsets with support greater than the minimum support, b) uses the frequent itemsets to generate the association rules. The algorithm finds only those rules that have the patient class attribute as a consequent in the rule. All other rules are ignored by the algorithm, for the domain user is not interested in such rules.

Placing aforementioned constraints on the algorithm search pattern is quite usual in the field of medicine where clinicians are not interested in finding all the associations in the datasets. For example, in hepatitis dataset, only the rules that show the association of the test findings with the patient class are meaningful for diagnosis purpose. The results presented in the next section have depicted that constraining the behavior of algorithm on-the-fly (i.e. interactively) helps clinicians to find attributes that determine the patient class.

INTERPRETATION OF RESULTS

Executing the algorithm with 15% support and 80% confidence has resulted into discovery of 48 association rules hidden in the hepatitis dataset. Figure 1.2 shows all the rules discovered by the algorithm with user specified support and confidence threshold. All these rules have class as target attribute. This experimentation gives rule induction method for prediction of patient class based upon the 19 recorded attributes of the hepatitis patient dataset. Upon close observation of the discovered rules, the attribute Anorexia, Proteome and Histology produced the best results. The method will predict which attributes contribute more to a person's chances of being attacked with hepatitis disease. This technique has been applied because of the ready availability of the subjects with some knowledge of the domain that can provide feedback on the explanations. The identification and interpretation of the discovered rules requires ample domain knowledge. For example, the rule {9, 15, 16, 18, 48, 55}-> 1 is translated as:

Male (X, "Yes") $\cap$ Anorexia (X, "Yes") $\cap$ Liver Big (X,"Yes") $\cap$ Spleen Palpable (X, "Yes") $\cap$ Protime30 (X,"Yes") $\cap$ Histology (X, "Yes") $\rightarrow$ Class (X, "DIE")

[Support = 15%, Confidence = 100%] [Rule 1]

(1) {15 48} -> {1}
100.0%

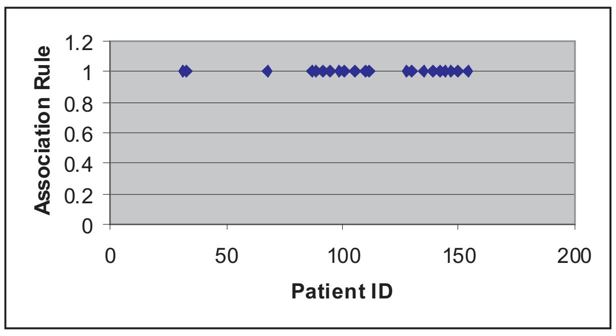
(2) {9 15 48} -> {1}
100.0%

(3) {12 15 48} -> {1}
100.0%

(4) {9 12 15 48} -> {1}
100.0%

(5) {15 16 48} -> {1}
100.0%

The rule induction method has the potential to use retrieved cases for prediction. A close look at each attribute in the rule points out the following findings:
The first part of the antecedent states that hepatitis is more common in males than in females.
Anorexia (state of loss of appetite) is a persistent problem with many chronic or serious diseases. The next part of the rule states that patients with acute hepatitis will suffer from anorexia.
The rule suggests that big liver and spleen disease virus (BLSV) is closely related to hepatitis virus.



The reference range for prothrombin time (a measure of ability of the blood to clot) is usually around 12–15 seconds. A prolonged or increased prothrombin time suggests that the blood under test is taking too long to form a clot which may be caused by conditions such as liver disease.

The rule indicates that the patients have gone through the histological1 reconstruction study of liver cell necrosis.

Once the association rules among various data attributes have been established, the important task that remains is to use these rules in biomedical research and patient treatment. In contrast to traditional data mining, involving domain expert in the analysis and interpretation of patient database can help in the proper diagnostic study of patient data. In the present work, the researchers were interested in finding only those association rules that determine the patient class value "DIE". Further, it has been found that values of the attributes in antecedent of the rules indicate that the disease is chronic. Such a rule can help clinicians to deal with the disease in a more effective way, in that, if the test results of a patient indicate some similarity with these rules then necessary actions can be taken for the patient so as to cure the infection before reaching it at malignant level.

CONCLUSION

Association rule mining as a data mining technique is very useful in the process of knowledge discovery in medical field, especially in the domain where patients' lab test reports have been electronically stored. Here association rule method is interactively implemented to predict the hepatitis patient's class. Such an experiment can give medical doctors a tool to quickly get some knowledge from the past patient's database and use them for handling future case. Understanding complex relationships that occur among patient's symptoms, diagnosis and behaviour is one of the most promising areas of data mining. The problem of identifying a patient's class is a major challenge among medical practitioners. Data mining techniques provide a tool to help them quickly make sense out of vast clinical databases.

HINTS

1.Histology is the study of the microscopic anatomy of cells and tissues of plants and animals. It is performed by examining a thin slice of tissue under a light microscope or electron microscope.

REFERENCES

1 R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital-to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/
2 M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/FLEXChip Signal Processor (MC68175/D), Motorola, 1996.“PDCA12-70 data sheet,” Opto Speed SA, Mezzovico, Switzerland.
3 A. Karnik, “Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP,” M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
4 J. Padhye, V. Firoiu, and D. Towsley, “A stochastic model of TCP Reno congestion avoidance and control,” Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
5 Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.