ISSN No. 0976-5697

Volume 4, No. 6, May 2013 (Special Issue)

International Journal of Advanced Research in Computer Science

REVIEW ARTICAL

Available Online at www.ijarcs.info

Survey on Data Mining Technique Using Decision tree For Hepatitis virus

Miss K. P. Lokhande
ME (CSE) 2nd Sem
G.H. Raisoni College of Engineering, Amravati
Kirti.lokhande15@gmail.com

Prof. Avinash. P. Wadhe
M-Tech (CSE)
G.H. Raisoni College of Engineering, Amravati
aviwadhe@gmail.com

Abstract: This work demonstrates the application of decision tree, as data mining tool, in the health care system. Data mining has the capability for classification, prediction, estimation, and pattern recognition by using health databases. Databases of health systems contain significant information for decision making. It could be properly revealed with the application of appropriate data mining techniques. Decision trees are employed for identifying valuable information in health databases. In this paper Decision tree as a data mining tools is used for predication the spread types of disease of hepatitis virus in reigns that high affect to people with different temperature and prevention of this disease by using rules that needed to predicate the diseases .

Keywords:- Decision Tree Modeling, Healthcare Data Mining Applications, Decision Rule Induction.

I. INTRODUCTION

Data mining can be defined as the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Alternatively, it can be defined as the process of data selection and exploration and building models using vast data stores to uncover previously unknown patterns. Data mining is not new it has been used intensively and extensively by financial institutions, for credit scoring and fraud detection; marketers, for direct marketing and cross-selling orup-selling; retailers, for market segmentation and store layout; and manufacturers, for quality control and maintenance scheduling. In healthcare, data mining is becoming increasingly popular, if not increasingly essential. Several factors have motivated the use of data mining applications in healthcare. The existence of medical insurance fraud and abuse, for example, has led many healthcare insurers to attempt to reduce their losses by using data mining tools to help them find and track offenders.3 Fraud detection using data mining applications is prevalent in the commercial world, for example, in the detection of fraudulent credit card transactions.

Recently, there have been reports of successful data mining applications in healthcare fraud and abuse detection. Another factor is that the huge amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining can improve decision-making by discovering patterns and trends in large amounts of complex data. 4 Such analysis has become increasingly essential as financial pressures have heightened the need for healthcare organizations to make decisions based on the analysis of clinical and financial data. Insights gained from data mining can influence cost, revenue, and operating efficiency while maintaining a high level of care.5 Healthcare organizations that perform data mining are better positioned to meet their long-term needs, Benko gives an illustration of a healthcare data mining application; and finally, highlighting the limitations of data mining and offering some future directions. Methods for classification and regression that have been redeveloped in the fields of pattern recognition, statistics, and mac, these are of particular interest for data mining since they utilize symbolic and interpretable representations. Symbolic solutions can pro-vide a high degree of insight into the decision bound-arise that exist in the data, and the logic underlying them. This aspect makes these predictive mining techniques particularly attractive in commercial and healthcare data mining applications. This paper present here a synopsis of some major state-of-the-art tree and rule mining methodologies [1,2,3].

II. HEALTHCARE DATA MINING APPLICATIONS

There is vast potential for data mining applications in. Generally, these can be grouped as the evaluation of treatment effectiveness; management of healthcare; customer relationship management; and detection of fraud and abuse. More specialized medical data mining, such as predictive medicine and analysis of DNA micro-arrays, lies outside the scope of this paper. Treatment effectiveness. Data mining applications can be developed to evaluate the effectiveness of medical treatments. By comparing and contrasting causes, symptoms, and courses of treatments, data mining can deliver an analysis of which courses of action prove effective. For example, the outcomes of patient groups treated with different drug regimens for the same disease or condition can be compared to determine which treatments work best and are most cost-effective[2,4]. Along this line, United HealthCare has mined its treatment record data to explore ways to cut costs and deliver better medicine. It also has developed clinical profiles to give physicians information about their practice patterns and to compare these with those of other physicians and peer-reviewed industry standards. Similarly, data mining can help identify successful standardized treatments for specific diseases.

In 1999, Florida Hospital launched the clinical best practices initiative with the goal of developing a standard path of care across all campuses, clinicians, and patient admissions [5].Good account of data mining applications at Florida Hospital also can be found in Gillespie[7] and Veletsos [6]. Other data mining applications related to treatments include associating the various side-effects of

© 2010, IJARCS All Rights Reserved

treatment, collating common symptoms to aid diagnosis, determining the most effective drug compounds for treating sub populations that respond differently from the mainstream population to certain drugs, and determining proactive steps that can reduce the risk of affliction. Healthcare management. To aid healthcare management, data mining applications can be developed to better identify and track chronic disease states and high-risk patients, design appropriate interventions, and reduce the number of hospital admissions and claims. For example, to develop better diagnosis and treatment protocols, the Arkansas Data Network looks at readmission and resource utilization and compares its data with current scientific literature to determine the best treatment options, thus using evidence to support medical care. Also, the Group Health Cooperative stratifies its patient populations by demographic characteristics and medical conditions to determine which groups use the most resources, enabling it to develop programs to help educate these populations and prevent or manage their conditions.1 Group Health Cooperative has been involved in several data mining efforts to give better healthcare at lower costs. In the Seton Medical Center, data mining is used to decrease patient length-of-stay, avoid clinical complications, develop best practices, improve patient outcomes, and provide information to physicians—all to maintain and improve the quality of healthcare [6,7]. Data mining can be used to analyze massive volume of data and statistics to search for patterns that might indicate an attack by bio -terror that effect to the people and in which reigns with different temperature to prevention from disease. [8]

III. KEY TECHNIQUES

A. *Association:*

Association (or relation) is probably the better known and most familiar and straightforward data mining technique. Here, you make a simple correlation between two or more items, often of the same type to identify patterns. For example, when tracking people's buying habits, you might identify that a customer always buys cream when they buy strawberries, and therefore suggest that the next time that they buy strawberries they might also want to buy cream.

B. *Classification:*

You can use classification to build up an idea of the type of customer, item, or object by describing multiple attributes to identify a particular class. For example, It can easily classify cars into different types (sedan, 4x4, convertible) by identifying different attributes (number of seats, car shape, driven wheels). Given a new car, might apply it into a particular class by comparing the attributes with our known definition. You can apply the same principles to customers, for example by classifying them by age and social group.

C. *Decision Tree Modeling:*

Decision trees are generated from training data (see table 1) in a top-down for specific direction. The initial state of a decision tree is the root node is assigned to the outlook. The node is split to three classes which are North Iraq, Middle Iraq and South Iraq. The process is recursively shared with viral hepatitis in the spit to types of viral hepatitis which are type A, type B and type C these classes called leaf nodes.

D. *Clustering:*

By examining one or more attributes or classes, you can group individual pieces of data together to form a structure opinion. At a simple level, clustering is using one or more attributes as your basis for identifying a cluster of correlating results. Clustering is useful to identify different information because it correlates with other examples so this approach can see where the similarities and ranges agree.

E. *Prediction:*

Prediction is a wide topic and runs from predicting the failure of components or machinery, to identifying fraud and even the prediction of company profits. Used in combination with the other data mining techniques, prediction involves analyzing trends, classification, pattern matching, and relation. By analyzing past events or instances, you can make a prediction about an event.

F. *Sequential patterns:*

Often used over longer-term data, sequential patterns are a useful method for identifying trends, or regular occurrences of similar events. For example, with customer data you can identify that customers buy a particular collection of products together at different times of the year. In a shopping basket application, you can use this information to automatically suggest that certain items be added to a basket based on their frequency and past purchasing history.

G. *Long-term (memory) processing:*

Within all of the core methods, there is often reason to record and learn from the information. In some techniques, it is entirely obvious. For example, with sequential patterns and predictive learning you look back at data from multiple sources and instances of information to build a pattern.

IV. IMPLEMENTATION

There is strong correlation between the data obtained from the Iraqi Health Ministry is, for example, that people infected with hepatitis virus differ archive a to prepare them in the table No. 1 in the three types of the disease as there is a relationship between the spread of the virus and the different between temperature and prevention of this disease. In table (1) show the incidence of the disease varies from one region to the other, where the liver infection in the northern region than other regions, as evident in the incidence of the disease in areas with high population density, as is evident in the maps of Iraq where the numbers of liver.

© 2010, IJARCS All Rights Reserved

Table 1: Dataset Number of people infected with hepatitis virus in the city

Out look	city	viral Hepatitis type A	Dangerous Class1	viral Hepatitis type B	Dangerous Class2	Dangerous Class2	Dangerous Class1
North Iraq	Dahuk	0	Less	37	middle	436	large
North Iraq	Arbl	0	Less	1	middle	231	large
North Iraq	Mousal	451	Less	229	middle	288	large
North Iraq	Karkok	0	Less	294	middle	1097	large
North Iraq	Sulaymaniyah	0	Less	0	middle	0	large
Middle Iraq	Salahedin	162	Less	195	middle	391	large
Middle Iraq	Diala	365	Less	65	middle	392	large
Middle Iraq	Baghdad	513	Less	1407	middle	3824	large
Middle Iraq	Anbar	45	Less	111	middle	402	large
Middle Iraq	Karbala	629	Less	88	middle	783	large
Middle Iraq	Babil	0	Less	315	middle	1542	large
Middle Iraq	Wasit	463	Less	134	middle	397	large
South Iraq	Qadisyah	186	Less	79	middle	62	large
South Iraq	Missan	313	Less	54	middle	63	large
Middle Iraq	Najf	0	Less	207	middle	43	large
South Iraq	Thiqar	0	Less	95	middle	04	large
South Iraq	Mutanna	80	Less	27	middle	043	large
South Iraq	Basra	0	Less	252	middle	314	large

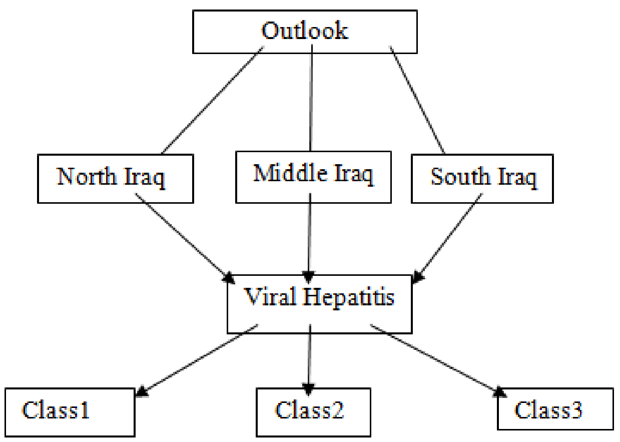

Figure1: The Decision Tree hepatitis virus

V. DECISION RULE INDUCTION

Decision rules, in disjunctive normal form (DNF), may be induced from training data in a bottom-up specific-to-general style, or in a top-down general-to-specific style, as in decision tree building. This section will highlight methodologies dealing with bottom-up specific-to-general approaches to rule induction. The initial state of a decision rule solution is indeed the collection of all individual instances or examples in a training data set, each of which may be thought of as a highly specialized decision rule. Most decision rule modeling systems employ a search process to evolve this set of highly specific and individual instances to more general rules. This search process is iterative, and usually terminates when rules can no longer be generalized, or some other alternate stopping criteria satisfied. As in the case of decision tree building, noise in the data may lead to over fitted decision rules, and various pruning mechanisms have been developed to deal with over fitted decision rule solutions. In figure (1) the tree have three leaf nodes. In decision tree, each leaf node represent a rule then have the following rule corresponding to the tree give.

Rule induction methods attempt to find a compact covering" rule set that completely partitions the examples into their correct classes. The covering set is found by heuristically searching for a single \best" rule that covers cases for only one class. Having found a \best" conjunctive rule for a class C, the rule is added to the rule set, and the cases satisfying it are removed from further consideration.

The process is repeated until no cases remain to be covered.

RULE1: In decision tree if it is North Iraq and type of hepatitis virus is A then Dangerous Is less.

RULE2: In decision tree if it is North Iraq and type of hepatitis virus is B then Dangerous Is middle Dangerous.

RULE3: In decision tree if it is North Iraq and type of hepatitis virus is C then Dangerous Is large

RULE 4: In decision tree if it is Middle Iraq and type of hepatitis virus is A then Dangerous is less.

© 2010, IJARCS All Rights Reserved

RULE 5: In decision tree if it is Middle Iraq and type of hepatitis virus is B then Dangerous Is Middle.

RULE 6: In decision tree if it is Middle Iraq and type of hepatitis virus is C then Dangerous Is large.

RULE7: In decision tree if it is South Iraq and type of hepatitis virus is A then Dangerous Is less.

VI. RESULT

From implementation of data mining Decision tree to the training set data it appear that.

a. Design Decision tree from dataset from table 1.
b. Design rules to predict the diseases and reigns.
c. The decision tree appears spread of disease and high population density.
d. The decision tree appears the types of disease and which high effect to the peoples.

VII. CONCLUSION

The rise in attention and focus on decision support solutions using data mining techniques has refueled a big interest in classification, particularly symbolic techniques. This paper has attempted to provide the reader with the key issues of decision tree and decision rule modeling techniques there is strong data as training data se obtained from the Iraq health Ministry. Data mining was success to solve the problem of effect hepatitis virus archive. Applied the concepts decision tree to real data and gained a working knowledge of data mining techniques. The rules in this design can predict which type of diseases the effect on the peoples and high affected also can appear the reigns the hepatitis virus spread. Therefore fundamental concepts of extracting knowledge from data should be a goal for .discovering important information.

VIII. REFERENCES

[1]. Milley, A., "Healthcare and data mining ".Health Management Technology, 21(8), 44-47, 2000.

[2]. Brewin, B. . "New health data net may help in fight against SARS". Computerworld, 37(17), 1, 59 , 2003 .

[3]. Rafalski, E. , " " Using data mining and data repository methods to identify marketing opportunities in healthcare. Journal of Consumer Marketing, 19(7), 607-613, 2002 .

[4]. Cios, K.J. & Moore, G.W. , "Uniqueness of medical data mining ". Artificial Intelligence in Medicine, 26(1), 1-24 , 2002

[5]. Arunk Pujori , " Data mining techniques universities press ", India private limited , 2008 .

[6]. Paddison, N. , " Index predicts individual service use ", Health Management Technology, 21(2), 14-17 , 2000

[7]. Schuerenberg, B.K. , " An information excavation ". Health Data Management, 11(6), 80-82 , 2003

[8]. www.ibm.com/legal/copytrade.shtml

Short Bio Data for the Author

**Miss.Kirti P. Lokhande:** Received the B.E and from SGBAU Amravati university.She is Currently pursuing M.E(CSE) in G.H Raisoni College of Engineering and Management,Amravati SGBAU Amravati university

**Prof. Avinash P. Wadhe:** Received the B.E and from SGBAU Amravati university and M-Tech (CSE) From G.H Raisoni College of Engineering, Nagpur (an Autonomous Institute). He is Currently an Assistant Professor with the G.H Raisoni College of Engineering and Management,Amravati SGBAU Amravati university. His research interest include Network Security , Data mining and Fuzzy system .He has contributed to more than 20 research paper. He had awarded with young investigator award in international conference.

© 2010, IJARCS All Rights Reserved

CONFERENCE PAPER
"A National Level Conference on Recent Trends in Information Technology and Technical Symposium" On 09th March 2013
Organized by
Dept. of IT, Jawaharlal Darda Inst. Of Eng. & Tech., Yavatmal (MS), India