

Research Article

STUDYING MEDICAL DATA MINING IN IRAN

Zahra Assarian Niassari and *Ahmad Faraahi

Department of Computer Engineering and Information Technology, Payame Noor University, Tehran,
Iran

*Author for Correspondence

ABSTRACT

In this paper, a review has been done to evaluate medical data mining. In order to study the characteristics of medical data mining an example of using data mining classification techniques were used.

Keywords: Data Mining, Data Mining, Medical, Data Mining Techniques

INTRODUCTION

In today's knowledge and control of data collection are plenty of different diseases. Medical centers with different purposes to collect data would pay for it. Research on the data and obtain useful results and patterns of disease Haas is one of the objectives of the data. The enormous volumes of data and confusion problem prevent it from achieving remarkable results (Mohammadi *et al.*, 2012).

Data Mining History of Medicine

Evans *et al.*, (1997) Stated that computer-based data mining techniques can be applied to a family history of clinical data, the algorithm is very accurate and clinically oriented model to detect hereditary diseases, such as colorectal cancer, the choice of data mining to evaluate HNPCC-related factors that , significant ($P < 0.05$) have been found. All final models HNPCC diagnosis is formulated independently by a clinician approval.

They concluded that the actions of family records from the previous analysis, the identification of specific inherited quite a negative response to a given family history of colon cancer and empirically the risk of enhanced. This can be used to select patients for DNA studies to facilitate the search for mutations. When mutations database contains parameters were a genetic disease, the result of an expert system that changes identified mutations in disease presentation. This information can greatly improve the efficiency of gene testing and automatic detection of hereditary syndromes is using data mining (Bharat *et al.*, 2010).

Medical Data Mining History in Iran

Saraei *et al.*, (2008) tried to natural disasters, according to the patient's first visit to the emergency information about accidents and property in the event of an accident, the possible factors influencing the relationship between the extractions of different types of accidents. In the end, it was discovered using a decision tree based on the rules, strategies to reduce child mortality by natural disasters, presented (Bharat *et al.*, 2002).

Khiyabani *et al.*, (2009) by applying data mining techniques on the records of patients with tuberculosis tried to concentrate on the early symptoms and test results of simple and quick responsive, timely and appropriate strategies for prevention and control of infection to others and offer healing disease (Jiawei and Micheline, 2006)

Rafiei (2009) classification studied the ways of dividing cancer, including analysis of Fisher's linear, weighted voting, Bayesian, Neural Network, Decision Tree, clustering, etc., and finally classification by support vector machine based on data as the best model introduced gene expression (Kay *et al.*, 2003).

Data Mining and Decision Tree Classification Method

One of the most important and most widely used data mining techniques of classification (classification) is a two-step method is used (Lee *et al.*, 1999). In the first step, a classifier based on a data set of labeled classes will be made clear at this stage is called learning. In the second stage classifier built on a set of data applied to the experimental data and the accuracy of the classifier tests. If accuracy is acceptable classifier can be applied on new data and determine their class. There are different methods for selecting test data. In cases where a large number of experimental data as separate data records from training data,

Research Article

i.e. data used to build the classifier used in these cases. The accuracy of a classifier on a test set as a percentage of records that have been correctly defined.

There are different options to choose the type of classifier. But one of the most important and yet most simple decision tree. Decision tree is a tree where each node is a flowchart-like internal (non-leaf node) in a test on attribute specifies. Each category displays the output of the test and each leaf node (terminal node) has a class label. Rectangular and oval leaves with known internal processes are displayed. No category (class), the amount of characters in the path from the root node of the tree is tested and a decision tree to x Suppose a data tuple to a leaf node in the cluster to determine tuple. Because of the simplicity and speed of decision trees to classify them in build and what is common. In general, decision trees, although accuracy is dependent on the successful use of the data used. Approach to building decision trees are generally top-down recursive divide and conquer is the input variable space tries to become the highest purity in the final processes. Decision trees are a special type of binary tree in which each node of the decision just two branches abroad (Roumani *et al.*, 2013).

Sample medical data mining applied to explore the relationship between cardiovascular risk factors.

The data used in this article, Isfahan Cardiovascular Research Center at the 1800 sampling plan, and test questions have been collected. Thus, the data set contains 1800 records, each record that is owned by one person. These data are available and include variables that the questionnaires were completed by the participants and the results of experiments SPSS file has been achieved in these cases. The questionnaire includes a variety of features such as nutrition, physical activity, smoking, and history of the people.

The technique is applied to classify the CART decision tree software product company Salford System is used for modeling. According to the recommended software to use cross validation options for the number of records less than 3,000 of these options were used for testing. Gini Splitting rule breaking in tough conditions as we choose.

The classification variables in Figure (2) show that the output is obtained CART.

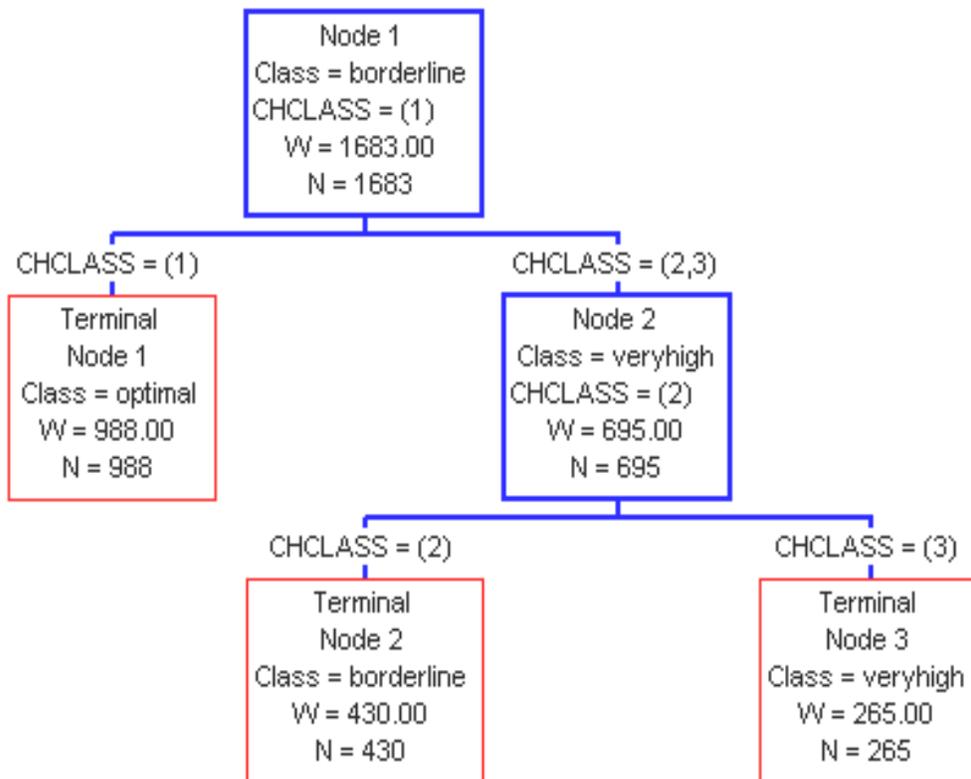


Figure 1: Decision tree made for LDL by selecting 27 predicting variables

Research Article

Variable	Score
CHCLASS	100.00
O12	15.19
BMI	11.58
APOB	11.19
TGCLASS	10.30
APOBA	5.81
APOA	2.38
S2	2.01

Figure 2: The most important variables for grouping variable target LDL

Observed in the classification of the target predicting variable cluster level is blood money. In other words, the change in LDL cholesterol is the most effects. As can be seen in the classification of other important variables, class variables are age, body mass index, Apo lipoprotein B, triglycerides level, of apolipoprotein B, Apo lipoprotein B, and the amount of smoking.

Conclusion

Using data mining to analyze medical data is a good opportunity to examine the relationships between variables. Despite the decision tree classification method gives results with acceptable accuracy in data mining, data collected in relation to the heart. Of appropriate data, the appropriate processing and data mining technique suitable for medical data gives good results. In this paper, examples of data mining medical data mining techniques, especially in the medical benefits were identified. Appropriate preprocessing and data mining technique suitable for medical data gives good results. By choosing clustering, association rules and those other clauses could be considered in future work and their results were compared with the results of this paper.

Since the amount of records to be used for data mining is more better results from the complete data set is useful in future investigations. Also, if the data collected during the data mining is a consistent means of data collected at intervals several times to be follows up on the data into the role of the time the can other applications will be considered.

REFERENCES

Bharat R, Krishnan S and Niculescu Radu S (2010). Data Mining for Improved Cardiac Care. *SIGKDD Explorations* 8(1).

Breault J, Goodall C and Fos P (2002). Data Mining a Diabetic Data Warehouse. *Artificial Intelligence in Medicine* 26 37-54.

Jiawei H and Micheline K (2006). *Data Mining, Concepts and Techniques*, second edition (Elsevier).

Kay Chen T, Yu Q, Heng CM and Tong Heng L (2003). Evolutionary computing for knowledge discovery in medical. *Artificial Intelligence in Medicine* 27(2) 129-154.

Lee ML, Lu H, Ling TW and KO YK (1999). Cleansing Data for Mining and Warehousing. *Proceedings of the 10th International Conference on Database and Expert Systems Applications (DEXA), Florence, Italy, August.*

Mohammadi Fard Noshin, Sarafzadegan N Zal, Gholamhossein Sadri, Gholam Hossein Malak Afzali, Shahnaz Shahrokhi Hamidreza Toluei, Masoud Pour Moghaddas, Masoumeh Sadeghi, Morteza Rafiei, Akbar Tavasoli, Kolishadi Roya Katayun Rabiei, Mansur Shahparian, Sedigheh Asgari, Gholam Naderi, Hassan Alikhassi, Changiz Tahereh, Alireza Yousefi and Hossein Samarian (No Date). The heart of the program, the program community-based interventions for the prevention and control of cardiovascular disease. *Journal of Research in Medical Sciences* 7(1).

Roumani YF, May JH, Strum DP and Vargus LG (2013). *Classifying Highly Imbalanced ICU Data* 16(2) 119-128.