

A Survey on Data Mining Tools and Techniques in Medical Field

D.Usha Rani MCA MPhil,
Assistant Professor, Department of Computer Science,
TBAK College for Women, Kilakarai.
Email: ushatharman84@gmail.com

ABSTRACT

In medical field the industry gathers massive amount of healthcare data which are not “mined” to find out unknown information. The medicinal industries come across with new treatments and medicine every day/every month/every year. The medicinal industries must deliver well conclusion and remedy to the patients to achieve worthy quality of service. Data Mining tools and techniques can be positively applied in many fields in different forms. Now Various Organizations using Data Mining as a powerful tool, to deal with the reasonable situation for data analysis. The primary purpose of this paper is to provide detail information about which tools and techniques are used to identify the accuracy level of various diseases. This data mining based prediction system reduces the human effects and cost effective one.

Keywords: **Data Mining: concepts-Tools and Techniques-Algorithms.**

1. Introduction

Data mining is also usually referred to as knowledge discovery from Data (KDD). The purpose of data mining is to mine useful information from huge databases or data warehouses. Now a day, Data Mining is becoming common in healthcare field because there is an essential of operational analytical methodology for detecting unidentified and valuable information in health data. In health industry, Data Mining offers numerous benefits such as recognition of the fraud in health insurance, availability of medical solution to the patients at lesser price, detection of bases of diseases and identification of medical treatment methods. Data mining algorithms useful in healthcare industry and shows an important role in estimate and finding of the diseases. There are a huge number of data mining applications are establish in the medical related areas such as Medical device industry, Pharmaceutical Industry and Hospital Management. To catch the valuable and unknown information from the database is the determination behind the application of data mining. The knowledge discovery is an interactive process, containing by developing an understanding of the application domain, choosing and making a data set, preprocessing, data transformation.

The data made by the health organizations is exact huge and difficult due to which it is hard to analyze the data in order to mark vital conclusion regarding patient health. This data covers details regarding hospitals, patients, medical claims, treatment cost and etc. So, there is an essential to make a powerful tool for analyzing and extracting significant information from this complex data. The analysis of health data expands the healthcare by improving the presentation of patient management jobs. The consequence of Data Mining technologies are to make available welfares to healthcare organization for

grouping the patients having related/similar type of diseases or health issues so that healthcare organization provides them active treatments. It can also valuable for predicting the how many days of stay of patients in hospital, for medical diagnosis and creating plan for active information system management. New and current technologies are used in medical field to improve the medical services in cost effective manner.

2. Classification of Data Mining System

Data mining systems can be categorized according to various criteria as given below

- 1) Type of data sources mined
- 2) Database involved
- 3) Kind of knowledge discovered
- 4) Mining techniques used

3. Process of Data Mining

Data mining process includes the following few steps

- 1) Data Cleaning** – It is used to remove noise and inconsistent data.
- 2) Data Integration** – It is used to combine multiple data sources.
- 3) Data Selection** –It is used to retrieve the relevant data from the database for analysis task.
- 4) Data Transformation** – It is used to transformed or consolidated data into particular appropriate form for mining by performing summary or aggregation operations.

5) Data Mining – Here the intelligent methods are applied in order to extract data patterns.

6) Pattern Evaluation – It is used to evaluate the data patterns.

7) Knowledge Presentation – Here the knowledge is represented.

The following figure 1 shows the data mining process.

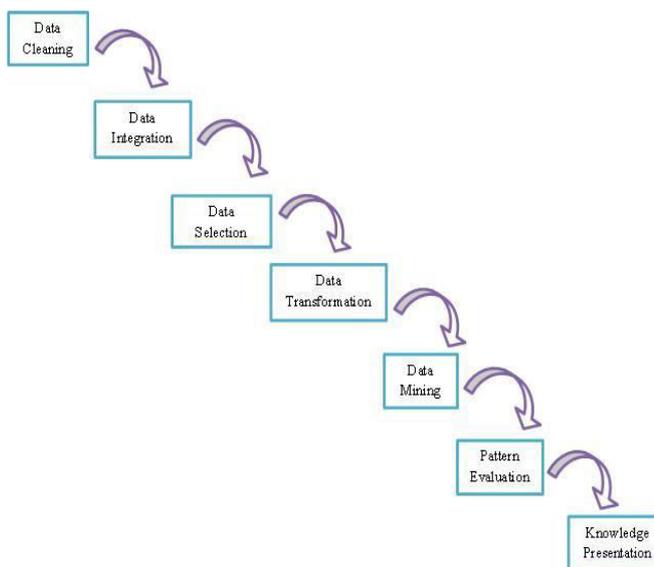


Figure 1: Process of Data Mining

4. Data Mining techniques

There are enormous number of data mining techniques have been evolving and using in data mining projects recently. Some of the data mining techniques are given below,

4.1 Association

Association is one of the top - well known data mining techniques. In association, a pattern is learned based on an association between items in the similar transaction. That's the purpose the association technique is also well-known as relation technique. The association technique is used in marketplace basket analysis to classify a set of products that customers regularly purchase together. Dealers are using association technique to investigation buyer's buying lifestyles. Based on ancient sale data, retailers might catch out that customers always buy jam when they buy breads, and, therefore, they can put jam and breads following to each other to save time for customer and make steps to growth sale.

4.2 Classification

Classification is a common data mining technique based on machine learning. Mostly, classification is used to categorize each item in a set of data into one of a predefined set of classes or groups. Classification method uses variety of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, make the software that can acquire how to classify the data items into groups. For instance, first apply classification in the application that "given all records of employees who left from the company; predict who will probably leave the company in a future period." In this case, we divide the records of employees into two groups that named "leave" and "stay". And then can ask our data mining software to classify the employees into separate groups.

4.3 Clustering

Clustering is a data mining technique that makes an expressive or valuable cluster of objects which have related or same characteristics using the automatic technique. The clustering technique describes the classes and puts objects in each class, while in the classification techniques, objects are allotted into predefined classes. To make the concept richer, can take book management in the library as an instance. In a library, there is a wide variety of books on different topics available. The challenge is how to hold those books in a way that readers can take several books on a particular topic without trouble. By using the clustering technique, can retain books that have some kinds of similarities in one cluster or one bookshelf and label it with a meaningful name. If readers need to take books in that topic, they would only have to go to that bookshelf instead of looking for the whole library.

4.4 Prediction

The prediction, as its name implied, is one of a data mining techniques that determine the association between independent variables and correlation between dependent and independent variables. For instance, the prediction analysis technique can be used in the sale to calculate income for the future if consider the sale is an independent variable, income could be a dependent variable. Then based on the ancient sale and earnings data, we can draw a fixed regression curve that is used for profit prediction.

4.5 Sequential Patterns

Sequential patterns analysis is one of data mining technique that pursues to determine or recognize associated patterns, fixed events or fashions in transaction data over a business period. In sales, with ancient transaction data, businesses can recognize a set of items that customers buy together different times in

a year. Then industries can use this information to mention customers buy it with better deals based on their purchasing regularity in the past.

4.6 Decision trees

A decision tree is one of the best common used data mining techniques because its model is easy to understand for users. In decision tree technique, the root of the decision tree is a simple question or condition that has compound answers. Each answer then leads to a group of questions or conditions that helps to determine the data so that can make the ultimate decision based on it. For example, use the following decision tree to determine whether a person has eligible for vote or not.



Figure 2: Decision Tree

Combine two or more of those data mining techniques composed to form an appropriate process that meets the business requirements.

5. Data Mining Tools

- Some of the data mining tools are given below,
- 1) Artificial Neural Networks (ANN),
 - 2) Rough Set Theory (RST),
 - 3) Statistical Package for the Social Sciences modeler (SPSS),
 - 4) K-means clustering
 - 5) Single Nucleotide Polymorphism (SNP)

Six best open source Data mining tools are given below,

- | | |
|------------------|------------------------------------|
| 1) Rapid miner | 2) Weka |
| 3) R-Programming | 4) Orange |
| 5) Knime | 6) Natural Language Toolkit (NLTK) |

6. Data Mining Applications in Healthcare

Data mining tools are used to predict the effective results from the data verified on healthcare problems. Different data mining tools are used to calculate the accuracy level in different healthcare problems. The following list of medical problems has been evaluated and estimated.

- | | |
|-------------------|-----------------|
| Heart Disease | Cancer |
| HIV/AIDS | Blood |
| Brain Cancer | Tuberculosis |
| Diabetes Mellitus | Kidney dialysis |
| Dengue | IVF |
| Hepatitis C | |

S No	Type of disease	Data Mining Tools	Data Mining Techniques	Algorithm	Accuracy Level in %
1	Heart Disease	ODND, NCC 2	Classification	Naïve	60
2	Cancer	WEKA	Classification	Rules .Decision Table	97.77
3	HIV /AIDS	WEKA 3.6	Classification and Association Rule Mining	J 48	81.88
4	Blood Bank Sector	WEKA	Classification	J 48	89.9
5	Brain Cancer	K-means clustering	Clustering	MAFI A	85
6	Tuberculosis	WEKA	Naïve Bayes Classifier	KNN	78
7	Diabetes Mellitus	ANN	Classification	C 4.5 algorithm	82.6
8	Kidney Dialysis	RST	Classification	Decision Making	75.97
9	Dengue	SPSS Modeler		C 5.0	80
10	In Vitro Fertilization (IVF)	ANN, RST	Classification		91
11	Hepatitis C	SNP	Information Gain	Decision Rule	73.20

Table 1: Tools and Techniques in Healthcare

The above table 1 show the various tools and techniques are used to find the accuracy level of various diseases.

7. Conclusion

The prediction of diseases using Data Mining applications is an inspiring task but it extremely reduces the human strength and increases the diagnostic accuracy. Developing well-organized data mining tools for an application could decrease the cost and time control in terms of human resources and capability. Discovering knowledge from the medical data is such a complicated job as the data found are noisy, irrelevant and massive too. In this scenario, data mining tools come in close in discovering of knowledge of the medical data and it is fairly interesting. It is detected from this table1 that a mixture of more than one data mining techniques than a single technique for diagnosing or predicting diseases in healthcare sector could produce additional encouraging results. The Table 1 displays the motivating results that data mining techniques in all the health care applications offer a more promising level of accuracy like 97.77% for cancer prediction and around 70 % for estimating the success rate of IVF treatment.

References

1. M. Durai raj, and V. Ranjani, "Data Mining Applications in Healthcare Sector: A Study", international journal of scientific & technology research volume 2, issue 10, October 2013 ISSN 2277-8616
2. HianChyeKoh and Gerald Tan, "Data Mining Applications in Healthcare journal of Healthcare Information Management" – Vol 19, No 2.
3. JayanthiRanjan, "Applications of data mining techniques in pharmaceutical industry", Journal of Theoretical and Applied Technology, (2007).
4. RubanD.Canlas Jr., MSIT, MBA, "Data mining in Healthcare: Current applications and issues".
5. K. Srinivas, B. Kavitha Rani and Dr. A. Govidhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks" International Journal on Computer Science and Engineering (2010).
6. D.Shobana, N.Uthra," The Data Mining Concepts and Techniques: A Survey" International Journal of Trend in Research and Development, Volume 2(6), ISSN 2394-9333
7. Hwang, H.G., Chang, I.C., Xingquan Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", ISBN 978- 1-59904-252, Hershey, New York, 2007
8. V.K.Deepa "Rapid Development of Applications in Data Mining" Proceedings of 2013 International

Conference on Green High Performance Computing
March 14-15, 2013, India

9. Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.

10. Han, J. & M. Kamber, and Data mining: concepts and techniques, San Francisco: Morgan Kaufman (2001).