A SURVEY ON MINING OF MEDICAL DATA IN DATA MINING : CHALLENGES AND ISSUES

S. Joyce[1], S.A. Sathya Prabha[2]

ABSTRACT

Medical Data Mining is acting as a vital role in Data Mining. The Medical data contains the Information of patient's record in a hospital. These data can be mined in order to get hidden patterns of data[1]. It helps to give a better treatment in medical. Using medical data mining framework the data is selected, preprocessed, mined and knowledge is evaluated[2]. This paper includes about medical data, fields of medical data framework, mining techniques, issues and challenges.

Keywords : Medical data mining, Association Rules, Classification, Clustering, Outliers.

I. INTRODUCTION

The Extraction of interesting patterns or knowledge is taken from a huge amount of data[1]. The data mining is interred burying penalizing area under discussion, which can be defined in many ways not only in medical field. It is grounded in artificial intelligence, databases, statistics, business, society, science, and engineering[3]. DM helps this information to build predictive models. The Role of DM in medicine (Healthcare) is enormous and multifaceted volumes of data are generated by healthcare activities; un-automated analysis has become impossible. DM can also create information including patient's details by identifying effective treatments and best practices[4].

II. MEDICAL DATA MINING :

In medicinal database have large quantities of data and information about patients and their medicinal terms. In history the past 20 years in with relational data bases has more dimensions and to access this databases using queries. In mid 1800's in London hit by infection diseases have two theories.

- Miasma theory (Bad Air Propagated Diseases)

- Germ theory (water-borne)

The Medical data mining is to discover the trends even when we do not realize the reasons to discover extraneous patterns that confuse than false belief.

In the field of medicinal mining lead to Question based answers, Anomaly based discovery, New Knowledge discovery, informed decisions, Probability measures, Predictive modeling, Decision support, improved health ,Personalized medicine.[5]

III. MEDICAL DATA MINING FRAMEWORK : (MDM)

The MDM framework includes medical data selection, medical data preprocessing, medical data mining and knowledge evaluation as shown in Fig.1.

[1]Asst. professor, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu - 641021.

[2]Asst. professor, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu - 641021.
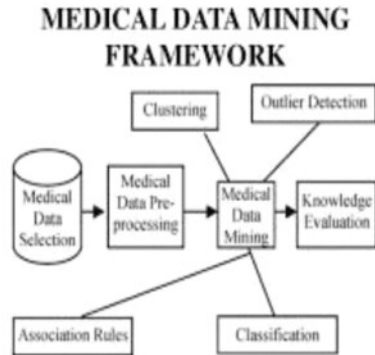
Figure 1. Medical Data Mining Framework.

Medical Data Selection :

The Medical data can be in the form of volume and complexity. The selected data has deprived mathematical categorization, canonical form and physician's interpretation.[6]

Medical Data Pre-processing :

In clinical databases the patient records consists of lab parameters, results of particular metrics. These are incomplete, noisy, inconsistent and temporal chronic diseases parameter that specific to task.[7]. If the medicinal data has no excellence then the mining results has no quality, because the data warehouse needs consistent integration of quality data. If people have domain knowledge they can handle incomplete, inconsistent or noisy data in medicinal domain.

Medical Data Mining :

Association Rules:

The Association rule defines (X)Y where X and Y are sets of items and T is set of transactions[8]. Here T is defined as patient's records and Items may be conditions of patient's record. The discovered association rules show hidden patterns in the mined dataset.

Table 1 : Training Set

PatientID	Conditions
P1	High LDL, Low LDL, High BMI, Heart failure
P2	Blood Pressure, Cancer, Fever
P3	Diabetes, Blood Pressure, Fever
P4	High LDL, Low HDL, Heart Failure
P5	Heart failure, High LDL, Low LDL, High BMI

For example the Association Rules of the above table {High LDL, Low HDL} ® {Heart Failure} People who have high LDL ("bad" cholesterol), low HDL ("good cholesterol") are at higher risk of heart failure.[13]

Classification :

A classification is the process of finding a patients record that describes the clinical information[10][18].The patients record are derived based on the analysis of set of training data[17]. For Example the classification model can be represented in various forms,

200

age(P1,"Youth")AND   disease(fever, "high")→class(P1,"A")
    age(P2,"Youth")AND disease(fever, "low")→class(P2,"B")
    age(P3,"middle-aged")→class(P3,"C")
    age(P4,"senior")→class(P4,"C")
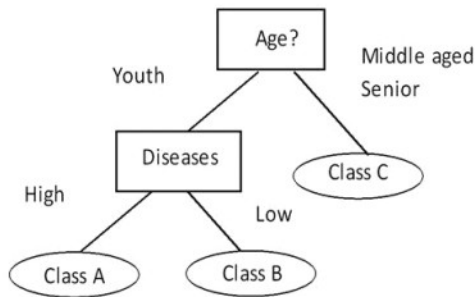
(a) IF-THEN Rules
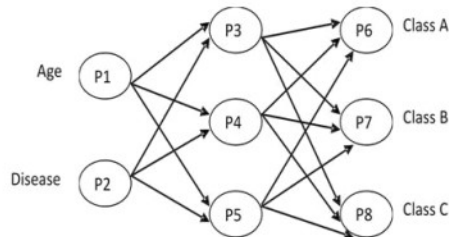
(b) Decision Trees

Clustering :

Clustering can be used to generate class labels for a group of patient's record[9]. In this technique the data items/set is partitioned in two or more clusters of similar patient records.

Outliers :

A data set contains objects that do not act in accord with the model of clinical data. These data objects are Outliers. Many data mining methods discards Outliers as exceptions. The analysis of Outlier data is referred to as Outlier analysis or anomaly mining [12].

Knowledge Evaluation :

The evaluation of knowledge in medical data mining is depends on the attributes of the clinical data values/measures [8].

Medical data mining challenges :

The medicinal data mining research and development problems can be classified as the following challenges.

Distributed Medical data :

The data to be minined is stored in distributed environment on heterogeneous platforms and it is impossible to bring all the data in a centralized place.

Distributed operation: In more data mining operation and algorithms will be available on grid in future. The distributed data mining for novel algorithm tools, grid services need to be developed.

Massive Medical data :

In medicinal data the development of algorithm for mining large, heavy and solid high dimensional data set is needed.

201

Complex Medical types :

Increased complex data sources like images, time series, mulit-relational, object data types and structures etc. Such data will require new methodologies and grid services.

Medical data, privacy, security and governance

In Grid based Medical data mining technology will need to address these issues.

Medical data mining issues :

The MDM development issues and problems as categorized as follows,

Table 2: Issues and Problems

S. No	MDM Issues	Problems
1	Patient Privacy	Aggregated data can be accomplished without patient identifiers[7]
2	Data quality	Cleaning the data is a teddies task to complete the research data warehouse. [11]
3	Standardized language	In health care – high quality data does not exists, lack of standardized data.[6][14]
4	Repeated Measures	Each repeated measure act as a new variable in data set. (mean, median, mode, standard deviation, large and frequency).
5	Missing values/variables	Elimination of the record, substitution of variable mean, substitution of Bayesian frequency. [16]
6	Overfitting data	Data are divided into training Vs testing sets in order to avoid over fitting.

VI. CONCLUSIONS

Data mining is extraction of patterns or knowledge from huge amount of data. medicinal is a hefty area, where data and information about patients are stored. The medical data mining framework include medical data, selection preprocessing, data mining and knowledge evaluation. In this paper the medical data mining challenges and issues are discussed. The privacy, security and misuse of information are the major problems. The data mining process is Systematic, it offers in medical the ability to discover hidden patterns in their data that can help them to understand patient's record.

REFERENCES

[1] Dhanya P Varghese & Tintu P B, "A Survey on Health Data Using Data Mining Techniques", International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 07, Oct-2015.

[2] Sheenal Patel and Hardik Patel "Survey of Data Mining Techniques Used in Healthcare Domain", International Journal of Information Sciences and Techniques (IJIST) Vol.6, No.1/2, March 2016

[3] Parvez Ahmad, Saqib Qamar, Syed Qasim Afser Rizvi, "Techniques of Data Mining In Healthcare: A Review", International Journal of Computer Applications (0975 – 8887) Volume 120 – No.15, June 2015

[4] Shomona Gracia Jacob , R Geetha Ramani, PhD, "Data Mining in Clinical Data Sets: A Review", International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 4– No.6, December 2012

[5] Mohammed Abdul Khaleel, Sateesh Kumar Pradhan, G.N.Dash, F. A. Mazarbhuiya," A Survey of Data Mining Techniques on Medical Data for Finding Temporally Frequent Diseases", International Journal of Advanced Research in Computer and Communication Engineering Vol.2, Issue 12, December 2013.

[6] Divya Tomar, Sonali Agarwal," A survey on Data Mining approaches for Healthcare", International Journal of Bio-Science and Bio-Technology Vol.5, No.5 (2013), pp. 241-266.

[7] Mary K. Obenshain, MAT,"Application of Data Mining Techniques to Healthcare Data", Infection Control and Hospital Epidemiology, Vol. 25, No. 8 (August 2004), pp. 690-695.

[8] Ruben D. Canlas Jr., "Data Mining in Healthcare: Current Applications and Issues".

[9] N.S.Nithya1, Dr.K.Duraiswamy, P.Gomathy," A Survey on Clustering Techniques in Medical Diagnosis", International Journal of Computer Science Trends and Technology (IJCST) – Volume1 Issue2, Nov-Dec 2013.

[10] M.A.Nishara Banu, B Gomathy, "Disease Predicting System Using Data Mining Techniques" International Journal of Technical Research and Applications e-ISSN: 2320-8163,

[11] Xin Xiao, Silvia Chiusano, "Analysis of Medical Treatments Using Data Mining Techniques", IEEE Intelligent Informatics Bulletin, December 2014 Vol.15 No.1

[12] Mohammed Abdul Khaleel, , Sateesh Kumar Pradham, "A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases", International Journal of Advanced Research in Computer Science and Software Engineering 3(8), August - 2013, pp. 149-153

[13] Dr.Shanmuga Priyaa,D., S.Kowsalya, 2015. A survey on Diagnosis Methods of Breast

203

Cancer using Mammography, International Journal of Engineering Technology Science and Research, 2 [4]

[14] Dr.Shanmuga Priyaa,D.,R.Sridevi, 2015. A Critical Study on Data Mining Techniques in Healthcare Dataset, International Research Journal of Engineering and Technology (IRJET), Vol 2, Issue 5, [10]

[15] Dr.Shanmuga Priyaa,D., S.Kowsalya , 2015. An Enhanced Approach for Preprocessing of Mammogram Images using Inverse Daubechies Wavelet Transform and Non-Linear Diffusion, International Journal of Computer Applications 130 (5), 6-12.

[16] Dr.Shanmuga Priyaa,D.,R.Sridevi, 2015. An Ensemble approach on Missing Value Handling in Hepatitis Disease Dataset, International Journal of Computer Applications 130 (17), 23-27.

[17] Dr.Shanmuga Priyaa,D.,K.Gomathi, 2016.Heart Disease Prediction Using Data Mining Classification, International Journal for Research in Applied Science & Engineering Technology (IJRASET),ISSN: 2321-9653, Vol 4, Issue II [60-63].

[18] Dr.Shanmuga Priyaa,D.,K.Gomathi, 2016. A Survey on Heart Disease Prediction Using Data Mining Techniques, International Journal of Multidisciplinary Educational Research, ISSN: 2277-7881; Impact Factor –4.527; IC alue:5.16; ISI Value:2.286, Volume 5, Issue 5(5),Pg: 56-63, May 2016.