International Journal of Research in Social Sciences
Vol. 9 Issue 2, February 2019,
ISSN: 2249-2496 Impact Factor: 7.081
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

DATA MINING TECHNIQUES IN HEALTH CARE

Monika*

Abstract

Data mining define as getting or "mining" the knowledge from large amount of data. As we know that mining of gold from rocks and sand is referred to as "Gold Mining" rather than rock and sand mining. So that "data mining" means "knowledge mining from data".

In the other words we can also define data mining such as "knowledge mining from databases, knowledge extraction, data/pattern analysis, and data dredging.

In this paper we discuss about data mining and its different type of techniques used in health care.

Keyword: Data mining, Database, Health Care

* **Assistant Professor, Ahir College Rewari, Haryana**

Introduction

Databases today can range in size into the terabytes-1,000,000,000,000 bytes of data. Within these masses of data lies hidden information of strategic importance. But when there are so many trees, how can you draw meaningful conclusion about the forest?

The newest answer is data mining, which is being used both to increase revenues and to reduce cost. Worldwide; organizations of all types are achieving measurable payoffs from this technology.

Data mining finds patterns and relationship in data by using sophisticated techniques to build model-abstract representation of reality. A good model is a useful guide to understanding your business and making decisions. Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses.

Many people treat data mining as a synonym for another popularly used term, \Knowledge Discovery in Databases", or KDD. Alternatively, others view data mining as simply an essential step in the process of knowledge discovery in databases. Knowledge discovery as a process is depicted in Figure 1.4, and consists of an iterative sequence of the following steps: data cleaning (to remove noise or irrelevant data), data integration (where multiple data sources may be combined)1 , data selection (where data relevant to the analysis task are retrieved from the database), data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)2 , data mining (an essential process where intelligent methods are applied in order to extract data patterns), pattern evaluation (to identify the truly interesting patterns representing knowledge based on some interestingness measures, and knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user)[1]
The data mining step may interact with the user or a knowledge base. The important patterns are presented to the user, and may be stored as new knowledge in the knowledge base. Note that according to this view, data mining is only one step in the entire process, albeit an essential one since it uncovers hidden patterns for evaluation. We agree that data mining is a knowledge

discovery process. However, in industry, in media, and in the database research milieu, the term \data mining" is becoming more popular than the longer term of knowledge discovery in databases". Therefore, in this book, we choose to use the term \data mining".[2] We adopt a broad view of data mining functionality: data mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories. Based on this view, the architecture of a typical data mining system may have the following major components: 1. Database, data warehouse, or other information repository. This is one or a set of databases, data warehouses, spread sheets, or other kinds of information repositories. Data cleaning and data integration techniques may be performed on the data. 2. Database or data warehouse server. The database or data warehouse server is responsible for fetching the relevant data, based on the user's data mining request. [3]3. Knowledge base. This is the domain knowledge that is used to guide the search, or evaluate the interestingness of resulting patterns. Such knowledge can include concept hierarchies, used to organize attributes or attribute values into different levels of abstraction. Knowledge such as user beliefs, which can be used to assess a pattern's interestingness based on its unexpectedness, may also be included. Other examples of domain knowledge are additional interestingness constraints or thresholds, and metadata (e.g., describing data from multiple heterogeneous sources). 4. Data mining engine.

DATA MINING TECHNIQUES

There are various type of data mining techniques such as association, classification and clustering are very useful in healthcare organization to enhance their capability for building appropriate conclusions for patient health from raw facts and figures .

1. Classification: Classification includes of two footsteps: - 1) Training and 2) Testing. Training makes a classification model on the basis of training data collected for generating classification rules. The IF-THEN prediction rule is highly popular in data mining; they signify facts at a high level of abstraction. The accuracy of classification model hinge on the degree to which classifying rules are true which is estimated by test data [9]. In health care domain classification can be made useful as “if Diabetic Family History=yes AND High Sugar Intake=yes THEN Diabetes Possibility=High”. Hat ice et al., to analyse skin diseases by using weighted KNN classifier [4].

2. Clustering is unalike from classification; it does not have predefined classes. A large database is broken into number of small subgroups which is known as clusters. It breaks the data based on similarities it have. Clustering algorithms discovers collections of the data such that objects in the same cluster are more identical to each other than other groups [5]. Tapia et al. examined the gene expression data with support of hierarchical clustering approach by using genetic algorithm [6].

3. Association also play very important role in the health care industry to discover the relationships between diseases, state of human health and the symptoms of disease. Ji et al., used association in order to learn uncommon casual relationships in Electronic health databases [10]. An integrated approach of using Association and Classification techniques also improved the capabilities of Data Mining. Sony et al., have used this integrated approach of association and classification for studying health care data. This integrated approach is useful for determining rules in the database and then by using these rules, an effective classifier is raised. The study made experiment on the data of heart patients and generate rules by weighted associative classifier [12]. Thus, Association also has an ample influence in the healthcare field to identify the relationships among various diseases, state of human health and the symptoms of disease.

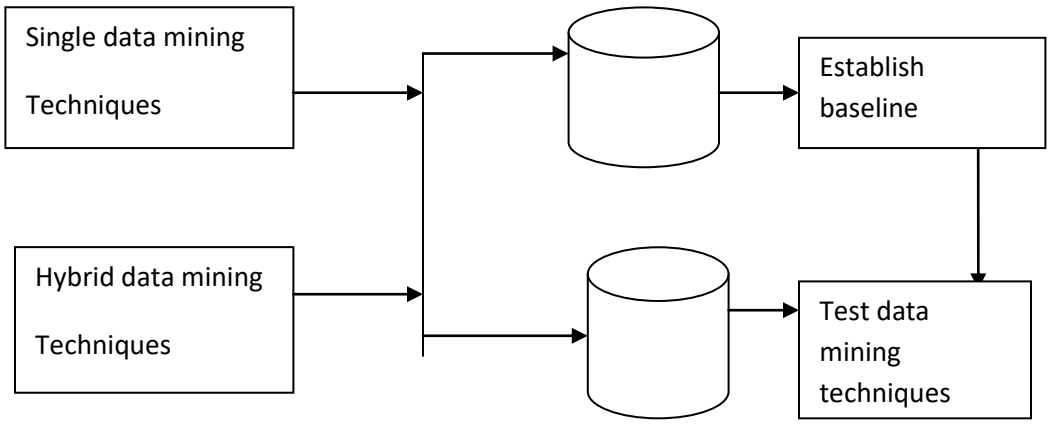

Figure: Proposed approach for data mining in health care

Conclusion

With the recent rapid rise in the quantity of biomedical data that is gathered by electronic means in critical care and the rampant availability of inexpensive and dependable computing equipment, many researchers has started, or are eager to start, exploring these data. In this paper we observe some data mining techniques that have been employed for medical data. As there is a voluminous record in this industry and because of this, it has become requisite to use data mining techniques to help in decision support and prediction in the field of Healthcare to identify the kind of disease. The medical data mining produces business intelligence which is useful for diagnosing of the disease. This paper throws light into data mining techniques that is used in health care.

REFERENCES

[1] C. Hattie & K. Matin, "A DIAGNOSTIC SOFTWARE TOOL FOR SKIN DISEASES WITHBASIC AND WEIGHTED K-NN", Innovations in Intelligent Systems and Applications (INISTA),2012.

[2] Dhanya P Varghese & Tints P B, "A SURVEY ON HEALTH DATA USING DATA MININGTECHNIQUES", International Research Journal of Engineering and Technology (IRJET), Volume:02 Issue: 07, Oct-2015.

[3] Doron Shalvi & Nicholas DeClaris, "AN UNSUPERVISED NEURAL NETWORK APPROACH TOMEDICAL DATA MINING TECHNIQUES", IEEE, 1998.

[4] Gustavo Santos-Garcia & Gonzalo Varela & Nuria Novoa & Marcelo F. Jimenez, "PREDICTION OFPOSTOPERATIVE MORBIDITY AFTER LUNG RESECTION USING AN ARTIFICIALNEURAL NETWORK ENSEMBLE", Artificial Intelligence in Medicine 30:61–69, 2004.

[5] Harleen Kaur & Siri Krishan Wasan, "EMPIRICAL STUDY ON APPLICATIONS OF DATAMINING TECHNIQUES IN HEALTHCARE", Journal of Computer Science 2 (2): 194-200, 2006.

[6] Hojin Moon & Hongshik Ahn & Ralph Kodell & Songjoon Baek & Chien- Ju Lin & James Chen,"ENSEMBLE METHODS FOR CLASSIFICATION OF PATIENTS FOR PERSONALIZED

MEDICINE WITH HIGH-DIMENSIONAL DATA". Artificial Intelligence in Medicine 41:197–207,2007.

[7] I. Curiac & G. Vasile & O. Banias & C. Volosencu & A. Albu, "BAYESIAN NETWORK MODELFOR DIAGNOSIS OF PSYCHIATRIC DISEASES", Proceedings of the ITI 2009 31st Int. Conf. onInformation Technology Interfaces, Cavtat, Croatia, 22-25 June-2009.

[8] Ilayaraja & T. Meyyappan, "MINING MEDICAL DATA TO IDENTIFY FREQUENT DISEASESUSING APRIORI ALGORITHM", Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, 21-22 February-2013.

[9] Illhoi Yoo & Patricia Alafaireet & Miroslav Marinov & Keila Pena-Hernandez & Rajitha Gopidi &Jia-Fu Chang & Lei Hua, "DATA MINING IN HEALTHCARE AND BIOMEDICINE: A SURVEYOF THE LITERATURE", Springer, May-2011.

[10] Jeong-Yon Shim & Lei Xu, "MEDICAL DATA MINING MODEL FOR ORIENTAL MEDICINE VIA BYY BINARY INDEPENDENT FACTOR ANALYSIS", IEEE.P1-4, 2003.

[11] J.J.Tapia & E. Morett & E. E. Vallejo, "A CLUSTERING GENETIC ALGORITHM FOR GENOMIC DATA MINING", Foundations of Computational Intelligence, Studies in ComputationalIntelligence, Volume:204, 2009.

[12] J.Yanqing & H.Ying & J.Tran & P.Dews & A.Mansour & R.Michael Massanari, "MININGINFREQUENT CAUSAL ASSOCIATIONS IN ELECTRONIC HEALTH DATABASES", 11th

IEEE International Conference on Data Mining Workshops, 2011.