

Data Mining and Life Science: A Survey

Dipti N. Punjani (Main Author)
Assistant Professor
National Computer College
Jamnagar
diptipunjani@gmail.com

Dr. Kishor Atkotiya (Corresponding Author:)
Professor
Department of Statistics
Saurashtra University- Rajkot
atkishor@yahoo.co.in

Abstract:- As we are into the age of digital information, the problem of data overload emerges so worryingly ahead. Our ability to analyze and understand immense datasets wrap extreme behind our ability together and stores the data. But a new age group of computational techniques and tools is required to support the extraction of useful knowledge from the rapidly increasing volumes of data. These techniques and tools are the focus of emerging fields of Knowledge Discovery in Databases (KDD) and also called data mining.

Data mining is highly noticeable in the fields like marketing, e-commerce or e-business or the fame of its use in KDD in other sectors or industries also. Among these sectors that are just discovering data mining are the fields of medicine and public health also. This research paper provides a survey of current technique of data mining/KDD for healthcare.

Keyword: Data Mining, Knowledge Discovery Database

I. Introduction

The purpose of data mining is to extract useful information from large databases or data warehouses. Data mining applications are used for different types of commercial and scientific surface (1). Scientific data mining differentiate itself in the sense that the nature of the datasets is often very different from traditional market driven data mining applications (2). Currently, different data mining algorithms applied in healthcare sector play a significant role in prediction and also diagnosis of different diseases. There are a different number of data mining techniques are establish in the medical related areas like Medical device industry, Pharmaceutical Industry and also Hospital Management. The data generated by the health sector is very vast and complex due to which it is difficult to analyze the data in order to make important decision regarding patient health. This data contains details regarding hospitals, patients, medical claims, treatment costs etc.

So, there is a need to generate a powerful tool for analyzing and extracting important information from this complex data. The examination or analysis of health data improves the healthcare by enhancing the performance of patient management tasks. The outcome of data mining technologies are provide different number of benefits of healthcare organizations like grouping the patients having similar type of diseases or health issues so that healthcare organization provides them effective treatments. It can also useful for predicting the number of days to stay of patients

in hospital, for medical diagnosis and making plan for effective information system management. Modern technologies are used in medical field to advance the medical services in cost effective manner. Data mining techniques are also used to scrutinize the various factors that are responsible for diseases for example types of foods, different working environment, education level, living conditions, availability of health care services, culture environmental and also agricultural factors (3).



Fig.1 Responsible Factors for Disease(4)

Medical data are characterized by their heterogeneity with respect to data type. These data may be noisy with erroneous or missing values. The records of millions of patients can be stored and computerized. However, there are other important issues such as ownership

and privacy related to these records. For example, cancer epidemiology is an important area of medical science where anatomic pathology reports can generate huge amounts of data to be mined for epidemiologic distribution of cancer (Cios & Moore,2000).

II. Literature Review

According to **HianChyeKoh and Gerald Tan**, data mining and its applications are useful in major areas of medical like treatment effectiveness, Management of healthcare, Detection of fraud and abuse and also Customer relationship management (1).

RazaAbidi(2001) has emphasized the involvement of knowledge management in the healthcare. In this paper ,the author contend that the operational effectiveness of a healthcare enterprise can be increased by using experimental knowledge to drive a group of packaged Strategic Healthcare Decision Support services (SHDS) derived from healthcare data and health organization knowledge bases. Specific types of SHDS include analysis of trends of admissions, treatments patterns, forecasting new diseases to evolve appropriate preventive measures, and also forecasting complications during the treatments (6).

JayanthiRanjan in this paper explained how data mining discovers and also extracts some useful patterns from the large data to find observable patterns. Through that paper, the author demonstrate the ability of data mining in improving the excellence of the decision making process in healthcare (7).

K.Srinivas et al., in this paper, discuss the potential use of different classification based data mining techniques such as Rule based decision tree, Naïve Bayes and also Artificial Neural Network to the massive data of healthcare (8).

ShwetaKharya also presented various approaches of data mining that have been utilized for breast cancer diagnosis and prognosis. In this paper, decision tree is found to be the best predictor with 93.62% accuracy on benchmark dataset and also on SEER data set (9).

According to **R.Vidya et al.**, (10) in this paper, the author has investigated different data mining techniques i.e. CART, RFT and K-means for the prediction of cervical cancer can be in different two stages that is Benign or Malignant or women with data mining algorithm with accuracy. During this study work, the accuracy percentage of CART is 83.87% with binary tree output.

To increase the correctness of the prediction level, RFT algorithm is used to predict cancer and it is classified as Benign or the accuracy level reached to the extent of

93.54% and also accuracy of RFT with K-means for cervical cancer prediction is improved to 96.77% while comparing these three.

Arvind Sharma et al., discussed data mining can easily use with important benefits to the blood bank sector. In this paper, they used J48 algorithm in WEKA tool. Classification rules performed well in the categorization of blood donors, whose accuracy rate reached 89.9% (11).

III. Data Mining

Data Mining came into existence in the middle of 1990's and appeared as a powerful tool that is suitable for fetching previously unknown patterns and useful information from large amounts of dataset. Various studies highlighted that data mining techniques help the data holder to analyze and discover unsuspected relationship among their data which in turn helpful for making decision (12). In general, Data Mining and Knowledge Discovery in Databases (KDD) are related terms, but many researchers assume that both the terms are different as Data Mining is one of the most important stages of the KDD process (13, 14). The knowledge discovery process are structured in various stages whereas the first stage is known as data selection where data is collected from various sources, the second stage is preprocessing of the selected data, the third stage is the transformation of the data into appropriate format for further processing, where as the fourth stage is Data Mining where suitable Data Mining technique is applied on the data for extracting valuable information and evaluation is the last and final stage as we can see in the below Figure. (13,15).

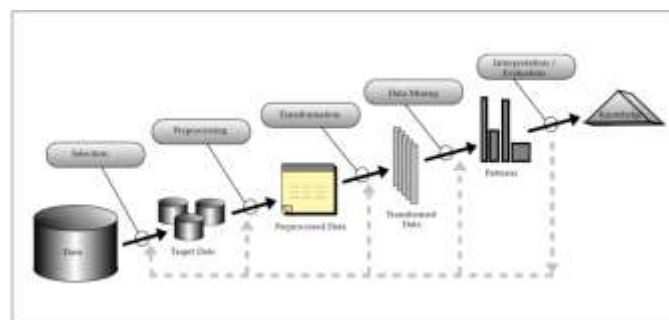


Fig 2: Knowledge Discovery of Database

Definition:-

Data Mining or knowledge discovery in database, as it is also known, is the not-trivial extraction of implicit, previously unknown and potentially useful information from the data. This includes a various number of technical approaches, such as clustering, data summarization, classification, finding dependency networks, analyzing changes, and detecting anomalies (16).

In current period various healthcare institutions are producing huge amounts of data which are difficult to handle. So, there is always need of powerful automated data mining tools for analysis and interpreting the useful information from this type of data. This information is very important for healthcare specialist to understand the reason of diseases and also for providing a better and cost effective treatment of patients. Data mining also offers useful information regarding healthcare which in turn to helpful for administration as well as medical decisions such as health care insurance policy, selections of different types of treatments, disease prediction, estimation of medical staff etc., (17-20).

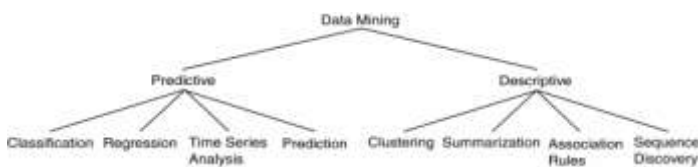


Fig 3: Data Mining Architecture

Data Mining Techniques:-

Data mining techniques are mainly divided into two categories:

- Predictive Techniques
- Descriptive Techniques



Data Mining Healthcare Applications:-

In current era various healthcare institutes are producing enormous amounts of data which become difficult to handle. So, there is a big need of powerful automated tools for analysis and also interpreting the useful information from this tremendous data. This type of information is very valuable and useful for any healthcare specialist to understand the reason of the diseases and also for providing different better and cost effective treatments to any patients. Data mining applications in healthcare can be grouped as the evaluation into broad categories (1, 7-11, 16). Following are several different applications of data mining in healthcare:

Healthcare Management:-

Data mining applications are used for better identification and also easy track any chronic disease into particular state and also high risk patients. Based on the complications of disease of any patient, hospital can easily set priorities of the patients so that they will get effective treatment in accurate manner and also in punctual time manner. This application is also useful to reduce the number of hospital admissions and also claims to assist healthcare management.

Treatment Effectiveness:-

This application can be developed to evaluate the effectiveness of medical treatments. According to K.J. Cios et al., by comparing and contrasting causes, symptoms, and also time schedules of treatments, data mining can deliver an analysis of which treatment is prove effective. Hospital can identify through the outcome of patient groups treated with different drug or treatment for the same disease or condition can be compared to determine which treatments work best and are most cost-effective (21). According to Hallick, data mining techniques are helpful to provide the information of patients regarding different diseases and also their prevention(22). Kolar, has documented that healthcare society uses data mining techniques for patient grouping (23).

Customer Relation Management:-

Customer relationship management is a core approach in managing interactions between commercial organizations- typically banks and retailers- and their customers. This application can be developed in the healthcare industry to determine the preferences, usage patterns, and current and potential needs of individuals to improve their level of satisfaction (24).

Decrease abuse and fraud:-

Healthcare insurer develops a model to detect the fraud and also abuse in the medical claims using data mining techniques. This model is useful for identifying the improper prescriptions, irregular or fake patterns in medical claims made by physicians, patients, hospitals etc (3).

IV. Conclusion

This paper explores the application of data mining in healthcare. Data mining provides benefit to the doctor, healthcare insurers, patients and also different healthcare organizations. Through data mining, doctor can easily recognize the effective cure, patients obtains cost effective treatments, healthcare organizations manages their patients

and also insurers discover any cases of fraud in medical claim.

Finally, the author hope of this paper can make a contribution to the data mining and healthcare industry and practice. It also is hoped that this paper can help all parties concerned in healthcare reap the benefits of healthcare data mining.

References

- [1] HianChyeKoh and Gerald Tan (2005), "Data Mining Application in Healthcare", Journal of Healthcare Information Management, Vol 19, No-2.
- [2] 2. M.Durairaj, V.Ranjani (Oct-2013), "Data Mining application in healthcare sector: A survey", International Journal of Scientific and Technology Research, Vol-2, Issue-10, pp.29-35 (ISSN 2277-8616)
- [3] Divya Tomar, Sonali Agarwal (2013), "A survey on Data Mining approaches for Healthcare" International Journal of Bio-Science and Bio-Technology, Vol-5, No-5, pp.241-266 (ISSN 2233-7849)
- [4] Dahlgren G & Whitehead M (1991), "Policies and strategies to promote social equity in health". Institute for Future Studies, Stockholm (Mimeo)
- [5] Cios, K. J., & Moore, G. W. (2000), "Medical Data Mining and Knowledge Discovery: An Overview". Heidelberg: Physica-Verlag.
- [6] Abidi, S.S.R (2001), "Knowledge Management in healthcare: towards 'knowledge-driven' decision -support services", International Journal of Medical Informatics 63, 5-18.
- [7] JayanthiRanjan (Dec-2007), "Application of data mining techniques in pharmaceutical industry", Journal of Theoretical and Applied Technology, Vol-3, No-4, pp. 61-67
- [8] K.Srinivas, B. Kavitha Rani and Dr. A. Govrghan (2010), "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks" International Journal on Computer Science and Engineering, Vol-2, No-2, pp.250-255.
- [9] ShwetaKharya (April-2012), "Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease", International Journal of Computer Science, Engineering and Information Technology(IJCSEIT), Vol-2, No-2.
- [10] R.Vidya, G.M.Nasira (August-2016), "Prediction of Cervical Cancer using Hybrid Induction Technique: A solution for Human Hereditary Disease Patterns" Indian Journal of Science and Technology, Vol-9(30), pp.1-10, ISSN 0974-6846.
- [11] Arvind Sharma and P.C.Gupta (2-Sep-2012), "Predicting the number of Blood Donors through their age and Blood Group by using Data Mining Tool" International Journal of Communication and Computer Technologies, Vol-1, No-6, pp.6-10.
- [12] D.Hand, H.Mannila and P. Smyth (2001), "principles of data mining" MIT.
- [13] U.Fayyad, G.Piatetsky-Shapiro and P.Smyth (1996), "The KDD process of extracting useful knowledge from volumes of data column" ACM, Vol.30, No-11, pp. 27-34.
- [14] J.Han and M.Kamber(2006), "Data Mining: Concept and techniques", 2nd edition, The Morgan Kaufmann Series.
- [15] U.Fayyad, G.Piatetsky-Shapiro and P.Smyth(1996), "From data mining to knowledge discovery in database" ACM, Vol.39, No-11, pp. 24-26.
- [16] Arun K Pujari (2006), "Data Mining Techniques", Universities (India) Press Private limited.
- [17] M. Silver, T. Sakara, H. C. Su, C. Herman, S. B. Dolins and M. J. O'shea (2001), "Case study: how to apply data mining techniques in a healthcare data warehouse", Healthc. Inf. Manage, vol. 15, no. 2, pp. 155-164.
- [18] P. R. Harper (2005), "A review and comparison of classification algorithms for medical decision making", Health Policy, vol. 71, pp. 315-331.
- [19] V. S. Stel, S. M. Pluijm, D. J. Deeg, J. H. Smit, L. M. Bouter and P. Lips (2003), "A classification tree for predicting recurrent falling in community-dwelling older persons", J. Am. Geriatr. Soc., vol. 51, pp. 1356-1364.
- [20] R. Bellazzi and B. Zupan (2008), "Predictive data mining in clinical medicine: current issues and guidelines", International Journal of Medical Informatics, vol. 77, pp. 81-97.
- [21] K. J. Cios, W. Pedrycz, and R. Swiniarsk (1998), "Data mining methods for knowledge discovery", Neural Networks, IEEE Transactions on, vol. 9, pp. 1533-1534.
- [22] J.N. Hallick (2001), "Analytics and the data warehouse", Health Management Technology, Vol.22, no-6, pp. 24-25.
- [23] H.R. Kolar (2001), "Caring for healthcare", Health Management Technology, Vol.22, no.4, (2001), pp. 46-47.
- [24] Biafore, S. (1999). Predictive solutions bring more power to decision makers. Health Management Technology, 20(10), 12-14.