

Data Mining Techniques in Health Care

Celeste Murnal¹, V. Deepika²

¹M. Tech. Student, Department of CSE, NHCE Bangalore, India

²Sr. Assistant Professor, Department of CSE, NHCE Bangalore, India

Abstract: Health care field is a highly intense field as it deals with the all-important process of delivering or technically sound and logical services for the treatment of patients. In this regard, it becomes very important to collect and analyze health related data from various sources comprehensively. It becomes convenient and easy for the health care providers to draw vital essence from such data and utilize them in improving health care delivery. The popularity gained by Data mining and warehousing in other fields such as research, teaching, manufacturing businesses, and marketing has also influenced the researchers and health care professionals to implement the system in the realm of health care, which is growing faster than ever. These techniques can reduce the redundancy and duplications in data compilation and produce more accurate, faster and reliable information. They have capabilities to understand hidden relationship patterns among the variables in health data. Learning about the disadvantages of these tools is equally important, in that, the privacy and quality concerns in relation to extracting data from Electronic Health Records could be threatening to the security of confidential data. However, there are reasons, which sound favorable enough to implement data mining and warehousing techniques. Therefore, this review focuses on determining the applicability of data mining and warehousing through its different techniques in the field of health care through its wide collection of advantages.

Keywords: Collection of medical data, genetic algorithm Diabetic Family History

1. Introduction

There has been a tremendous transformation in the past decades in the domain of application of electronics and computerization. Electronic systems have made tasks easier and beneficial in the process of raising the level of efficiency and performances of the employees. Health care is one such field, which employs electronic systems and research techniques in the analysis and interpretation of its large volumes of data. Health care organizations and institutions worldwide have converted their paper-based data collection and processing into electronic formats. Data sets produce enormous amounts of data, which yield invaluable pieces of information. This also poses a challenge to the data managers or holders in relation to maintenance of feasibility of data. There is a constant need of reducing unnecessary wastage and increasing overall quality and operational efficiency of an organization, and this need has paved way for concepts as 'Data mining and warehousing which are proven significant. Data mining is defined as "a process of nontrivial extraction of implicit, previously unknown and potentially useful information from the data stored in a

database". [1]. There are huge databases from which raw or useless data are extracted through the process of Data Mining. Health care field specialists are one such group of beneficiaries from such mined data as they have learnt over the time how to extract vital data from big chunks of it. [2]. For an example, Canada is one of those countries that invest large amounts of money for the purpose of Data mining; this in turn is a proving to be a financial challenge for the government. According to an article in the policy magazine, estimation shows that by 2020, Canada will be spending a quarter of a trillion dollars on healthcare. According to an article in the Policy magazine, Canada will be spending a quarter of a trillion dollars on healthcare by 2020. The analyses can be helpline for financial, clinical and administrative issues of Canada's healthcare reforms; this can be equally helpful in other countries as well. The concepts and applications of Data mining and Data warehousing can to the maximum extent serve the purpose in relation to health care data informatics. The techniques such as classification, clustering, Bayesian networks association, neural network, and genetic algorithms among others are instrumental in the process of health care data collection and processing. [3].

2. Data mining in health care and its applications

Delivery of patient care services give birth to Health care data. These data include patients' demographical details, past and present health conditions, medical and surgical management and other related details. Patients themselves are deemed as the primary sources of their health data, whereas, other sources include hospital records, vital registers, publications, all of which are secondary sources. Hence, mining of medical data includes legal and privacy concerns. Collection of medical data, its analysis and interpretation is not an easy task since the sources are not always accurate to fetch data from. There could be changes in the way and accuracy of how one communicates his/her health related data. Accuracy and completeness of this data is mandatory for mining to take place appropriately. This fact has to be laid emphasis upon, since it provides base and rationale for the health care providers to play the role of decision makers in diagnosis, treatment and follow up of health related problems. Data mining thus helps in obtaining data in non-invasive, painless and effective manner. [4].

3. Data mining techniques

Data mining techniques include association, classification and clustering and others to extract data from root level and utilize them in processing further and interpreting the data thereby.

A. Classification

Classification comprises of two footsteps: 1. Training 2. Testing. A classification model is used to build Training on the basis of data collected for generating classification rules. One highly popular rule in data mining is the ‘IF-THEN’ prediction rule. 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 [5]. In health care domain classification can be made useful as “if Diabetic Family History=yes and High Sugar Intake=yes, then Diabetes Possibility=High”, to analyze skin diseases by using weighted KNN, as suggested by Hatice et al. [6].

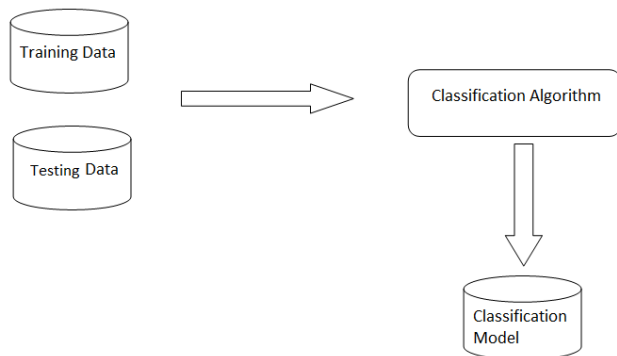


Fig. 1. Step in classification algorithm

B. Clustering

Clustering is different from classification; there are no predefined classes in it. Clusters are formed from subgroups divided from a large database. Data is based on similarities it has. Data collections are discovered from Clustering algorithms, this makes objects in the same cluster more identical to each other than the other groups. Tapia et al. used genetic algorithm to examine the gene expression data with support of hierarchical clustering approach. [7].

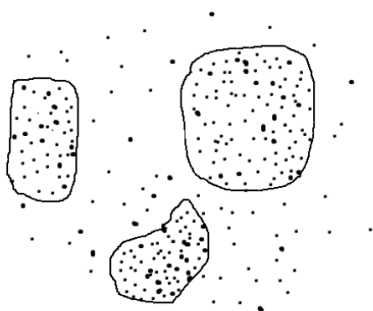


Fig. 2. Clustering of patients

C. Association

There is a great impact of association in the health care industry in discovering the relationships between diseases, state of human health, and the disease symptoms. Soni et al, have used this integrated approach of association and classification for studying health care data. Data base rules can be determined by this integrated approach and thereby these rules can be used to raise an effective classifier. This study made an experiment on the data of heart patients and generated rules by weighted associative classifier [8]. Therefore, there is significant influence of Association on healthcare field to identify the relationships among various diseases.

4. Other techniques in health care

Mining medical data for various routine purposes and tasks is done with the help of various data mining techniques. These techniques are used for data selection as they provide easy access to most required data at crucial times. Health care is one such important field that emphasizes on the importance of data collection from wide array of sources by employing the following techniques.

A. Neural networks

Neural Networks came into existence and was developed in the 20th century [9]. Decision trees and support vector were previously used before neural networks. These machine supported neural networks trees were earlier the greatest cataloguing algorithm. These are used in health care for drug developments, image interpretations, and clinical diagnoses.

B. Decision tree

A test or a situation on a data article is illustrated by each non-ultimate join. These fatal and non-fatal join is illustrated on a decision tree. Instances are classified by cataloguing the points from joining of fatal-non fatal points on a decision tree. Decision trees are used commonly for scheming qualified odds in action. They are used in health care when choices and outcomes of treatment are uncertain, and when such choices and outcomes are significant (wellness, sickness, or death). They also enable decision makers to determine and decision ways to reduce risks associated with certain procedures and treatment modalities.

C. Fuzzy sets

It is one of the best methods employed in linguistics, decision making and clustering. It permits the gradual assessment of membership of elements in a set, which is described with a membership value denoted by real unit interval (0, 1). Fuzzy sets are used in a wide range of areas or domains where information is incomplete or imprecise, which includes Health care. The fuzzy sets help to estimate the pain in under intensive procedures such as anesthesia, and are also used to predict response of patient to treatment, to improve decision making in radiation therapy, and control hypertension during anesthesia etc.

Table 1
 Following list consist of uses of data mining in health care

S. No	Uses	Advantages	Methodology
1.	Early detection and prevention of diseases	Disease patterns, relationship among variables of diseases, and outliers can be identified by Medical practitioners Medical practitioners can identify patterns and outliers	This is done by using data mining method rather than just by looking at the tabulated data
2.	Cost-effective and painless diagnosis	Procedures to detect some life-threatening diseases such as cancer of cervix or ovaries are painful and highly expensive. Medical practitioners first use data mining algorithms.	By using data mining algorithms ,medical experts can determine and decide whether to have a biopsy to detect cervical cancer or not.
3.	Adverse drug events	Data mining enables the health care team to discover side and adverse effects of drugs used for the prevention and treatment of various diseases and disorders adverse effects.	Collection of wide range of data from various sources such as health status of the patient, drug nature and bio-chemical factors, pharmaceutical data, drug prescriptions etc. enables the experts to detect drug
4.	Hosting of safety issues and prevention of hospital errors	Plenty of vital and useful information about safety and errors can be gathered using data mining. Many unknown medical errors can be revealed through data pre-processing	This is done by applying data mining methods on their existing databases.
5.	Medical research	Several medical and surgical interventions are discovered constantly across the global through the application of data mining and warehousing.	This is done by using data mining. Collection of data related to hospitalization, surgical data, and outcomes, and patient feedbacks from the health care sectors enables the decision makers of the field in this regard.
6.	Public/Community health recommendations	Health care in public health care needs recommendation areas such as sterilization effects, health education impacts, health programs and vaccination results and positive outcomes, which is achieved through data mining.	This is done by gathering information from various sources in the rural and urban communities such as health registers, antenatal and pregnancy records, and vaccination details from health set-ups in public health sector.

D. Bayesian Networks

Acquaintance of a doubtful sphere is illustrated with the help of an explicit sort of set-up known as Bayesian network. For an instance, these networks can represent probabilistic relationships between disease and symptoms, thereby computing the probability of various diseases.

E. Genetic algorithm

They are commonly used to generate high quality and efficient solutions to search solutions of complex problems in the field of health care, genetics and genomics on the basis of bio-inspired operators such as mutation, crossover and selection.

5. Conclusion

As understood through this review it is evident that the enormous of data collected constantly from various health care sources is not easy to be compiled and processed. Data mining and warehousing play pivotal role in this regard. Operational

efficiency and quality of health care services provided are determined by the information gathered through different sources. Optimum utilization of these data can cut extravagant expenses incurred on health care if data mining is applied. Therefore, data mining and its techniques have received a lot of importance in health care field due to their ability to draw information from data. These effects have resulted in innovations of data mining algorithms, and new versions and modules are developed to provide accurate and comprehensive data. Health care field has to recognize these concepts, which have made disease prevention, diagnosis, treatment, and research activities more effective and impactful. [1]

References

[1] D. Tomar and S. Agarwal, 'A survey on Data Mining approaches for Healthcare', Int J. Bio-Sci. Bio-Technol., vol. 5, no.5, pp. 241-266, 2013.
 [2] Q. Lu. and L. Getoor, "Link-based classification," in Proc. 20th Int. Conf. Mach. Learning, Washington, D.C., USA, 2003, pp. 496-503.
 [3] J. Natale, 'Leveraging Technology to Revolutionize Canadian Health Care', Policy: Canadian Politics and Public Policy, vol. 2, no.6, pp. 27-30, Dec-2014

- [4] Ichise, R, and Numao Learning M., First-order rules to handle medical data, NII Journal 2:9-14, 2001.
- [5] Jia-Fu Chang & Lei-Hua, "Data mining in Healthcare and Biomedicine: A Survey of the literature", Springer, MAY-2011.
- [6] C. Hattice & K. Metin, "A Diagnostic Software tool for skin diseases with basic and weighted K-NN", Innovations in Intelligent systems and Applications (INISTA), 2012.
- [7] Geomic Data Mining", Foundations of Computational Intelligence, studies in Computational Intelligence, Volume: 204, 2009.
- [8] S. Soni & O.P Vyas, "Using Associative Classifiers for Predictive Analysis in HealthCare Data Mining", International Journal of Computer Applications. Volume: 04, No: 05, July-2010.
- [9] Arvind Sharma and PC Gupta, Predicting the number of blood donors through their age and blood group by using data mining tool. International Journal of communication and computer technologies. 1(6): 6-10, 2012.