

A study based on Healthcare Systems using Data Mining

R.Deepa, E.Saranya

Author 1: Assistant Professor, PSG College of Arts and Science, Coimbatore, India

Author2: Research Scholar, PSG College of Arts and Science, Coimbatore, India

ABSTRACT

Data Mining is one of the most motivating areas of research that is become increasingly popular in health care information system. Data Mining plays an important role for uncovering new trends in healthcare organization which in turn helpful for all the parties associated with this field. This survey explores the utility of various Data Mining techniques such as classification, clustering, association, regression in health care system. In this paper, we present a brief introduction of these techniques and their advantages and disadvantages. This survey also highlights applications, challenges and future issues of Data Mining in healthcare information system. Recommendation regarding the suitable choice of available Data Mining technique is also discussed in this paper. Data mining has an infinite potential to utilize healthcare data more efficiently and effectually to predict different kind of disease. This paper features various Data Mining techniques such as classification, clustering, association and also highlights related analyzed and predict human disease.

Keywords: *Data Mining, Classification, Clustering, Association, Healthcare system.*

1.Introduction

Data mining can be defined as the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Alternatively, it can be defined as the process of data selection and exploration and building models using massive data stores to uncover previously unknown patterns. Data mining is an analytic process designed to explore

large amounts of data in search of consistent patterns and/or systematic relationships between variables,

and then to validate the findings by applying the detected patterns to new subsets of data.

[1]. The data generated by the health organizations 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 cost *etc.* So, there is a need to generate a powerful tool for analyzing and extracting important information from this complex data. The analysis health data improves the healthcare by enhancing the performance of patient management tasks. The outcome of Data Mining technologies are to provide benefits to healthcare organization for grouping the patients having similar type of diseases or health issues so that healthcare organization provides them effective treatments. [2] manufacturers for quality control and maintenance scheduling and it has been used in hospital care as well. The existence of medical insurance fraud and abuse, for example has led many healthcare insurers to attempt to reduce their losses by using data mining tools, the application has helped to help them find and track offenders[2]. However fraud detection using data mining applications is prevailing in the commercial world for detection of fraudulent credit card transactions and fraudulent banking activities Data mining applications.

The aims of quality healthcare services are:

- providing safe healthcare treatments
- using scientific medical knowledge to provide healthcare services to everyone
- providing various healthcare treatments based on the patient's needs, symptoms and preferences

- minimizing the time to wait for the medical treatment
- minimizing the delay time in providing medical treatment.
- Health determinants
- Inputs to the health system and related processes (e.g., health infrastructure, human and financial resources, equipment, policy, and organization)
- Health outcomes (e.g., mortality, morbidity, disability, well-being, and health status)

2. DATA MINING

Data mining is the process of discovering interesting knowledge,[3] such as patterns, associations, changes, anomalies and significant, from large amount of data stored in database. Data warehouse or other information repositories. The wide availability of huge amounts of data in electronic forms ,and the imminent need for Turing such data into useful information and knowledge for broad application including market analysis business management, and decision support, data mining has attracted a great deal of attention in information system[3].

Data mining consists of five major elements:

1. Extract, transform, and load transaction data onto the data warehouse system.
2. Store and manage the data in a multidimensional database system.
3. Provide data access to business analysts information technology professionals.
4. Analyze the data by application software.
5. Present the data in a useful format, such as a graph or table.

3. Data mining application areas

Data mining is driven in part by new applications which require new capabilities that are not currently being supplied by today's technology. these new applications can be naturally into two broad categories.

- Business and e-commerce.
- Scientific ,engineering and health care data

3.1 Data Mining Models

Generally, there are two kinds of data mining models: predictive model and descriptive model [4].The predictive model often apply supervised learning functions to predict unknown or future values of other variables of interest. The descriptive model on the other hand, often apply the unsupervised learning functions in finding patterns describing the data that can be interpreted humans [4]. The data mining models identified among the papers reviewed are summarized The predictive models are more commonly used in the healthcare.

The two different models

1. Predictive
2. Descriptive

3.2. Data Mining Tasks

Usually, the implementation of a model is made by a task. For instance, clustering [5], association rules [6], correlation analysis, are often used for descriptive models. While classification [7], are used for predictive model.

The three different models

1. Classification
2. Association rules
3. Clustering

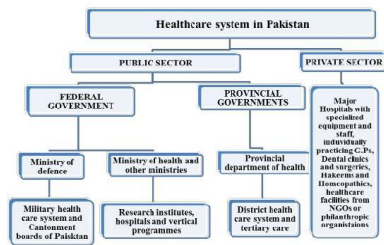
4. Data mining application in healthcare

Healthcare[8] industry today generates large amount of complex data about patients, hospital resource, disease diagnosis, electronic patient records, medical devices etc. larger amounts of data are a key recourse to be processed and analyzed for knowledge extraction that enables support for cost-saving and decision making. Data mining applications in healthcare can be grouped as the evaluation into broad categories[8].

4.1 Treatment effectiveness

Data mining application [8] can develop to evaluate the effectiveness of medical treatments. Data mining can deliver an analysis of which course of action proves effective by comparing and contrasting causes, symptoms, and courses of treatments.

HEALTHCARE SYSTEM



4.2 Healthcare management

Data mining application can develop to better identify and track chronic disease states and high-risk patients, design appropriate interventions, and reduce that number of hospital admissions and claims to aid healthcare management. Data mining used to analyze massive volumes of data and statistics to search for patterns that might indicate an attack by bio-terrorists.

4.3 Customer relationship management

Customer[9] relationship management is a core approach to managing interactions between commercial organizations typically banks and retailers—and their customer, it is no less important in a healthcare context. Customer interactions may occur through call centers, physicians, offices, billing, department, inpatient settings, and ambulatory care settings[9].

4.4 Health Informatics

Health informatics involves systems such as electronic health records (EHR), health information exchange (HIE) standards and portable medical data collection devices. Health Informatics is a term describing the acquiring, storing, retrieving and using of healthcare information to foster better collaboration among a patient's various healthcare

providers. Health Informatics plays a critical role in the push toward healthcare.

5. Health care algorithm system

5.1 Classification

Classification is the discovery of a predictive learning function that classifies a data item into one of several predefined classes [10]. The related work in classification will be discussed in the following subsections.

5.1.1 Statistical

The MTS algorithm is being extensively applied in multivariable statistical analysis. The Mahalanobis distance (MD) is used to build statistical judgments to distinguish one group from another and the Mahalanobis space (MS) is used to represent the degree of abnormality of observations from the known reference group. In the statistical classifiers, the authors Su et al. [20], have used the Mahalanobis Taguchi System (MTS) to design the prediction model for pressure ulcers. The class imbalance problems are very much prevalent in the healthcare datasets. The test conducted using these algorithms were done in four phases with scaled datasets ranging from 14 to 8, 5, and 2 accordingly. The results obtained in the paper [20] show that the measurement scale for this algorithm has good performance based on the huge difference between the normal and abnormal examples. Being an algorithm which is suitable for scaling the MTS proves to have better sensitivity and g-means values in the testing stage. The MTS has enhanced performance in terms of sensitivity.

5.1.2 K-Nearest Neighbor

Authors García-Laencina, Armañanzas et al. [11], Nahant al. have used the k-nearest neighbour in their respective predictive models. The k-nearest neighbor is an instance based classifier method. The parameter units consist of samples that are used in the method and this algorithm then assumes that all instances relate to the points in the n -dimensional space R^N . The algorithm is very expedient as the information in the training data is never lost.

5.2.3. Support Vector

The support vector method (SVM) is proven to be advantageous in handling classification tasks

with excellent generalization performance. The method seeks to minimize the upper bound of the Generalization error based on the structural risk minimization principle. The SVM training is equivalent to solve a linear constrained quadratic programming problem [10]. The method is very commonly used in have used the method in their model in medical diagnoses. Some of the authors have used the SVM method for comparative study purpose.

5.2 Clustering

The clustering is a common descriptive task in which one seeks to identify a finite set of categories or clusters to describe the data [10]. Rui Volusia had used the vector quantization method in clustering approach in predicting the readmissions in intensive medicine. The algorithms used in the vector quantization method are k-means, k-medicos and x-means. The k-means obtained the best results while x-means obtained fair results while the k-mediods obtained the worst results. From the results the work by these researchers provide a useful result in helping to characterize the different types of patients

5.3. Association

Association also has great impact in the health care industry to discover the relationships between diseases, state of human health and the symptoms of disease. used association in order to have used this integrated approach of association and classification for studying healthcare data. This integrated approach is useful for determining rules in the database and then busing these rules, an effective classifier is raised. The study made experiment on the data of heart patients and generate rules by weighted associative classifier.

6. Advantages and Disadvantages of Data Mining In Health Care:

Advantages:

Data mining can be advantageous as it can help

- Healthcare insurers detect fraud and abuse.
- Physicians identify effective treatments and best practices.

Patients receive better and more affordable healthcare One of the key advantages of using data mining is their speed in working with large data sets. Generation of quicker report and faster analysis that can increase operational efficiency and reduce operating cost.

- Data Mining can extract predictive information from large databases which is a very important feature of Data Mining.

Disadvantages:

- Heterogeneity of medical data Volume and complexity.
- Physician's interpretation Poor mathematical categorization
- Ethical, Legal and Social Issues.
- Data Ownership Lawsuits.
- Privacy and Security of Human Data Administrative.
- Privacy issues.
- Security issues.

7. Limitations and Challenges Of Data Mining Algorithms In Healthcare Services.

Medical diagnosis is considered as a significant yet obscure task that needs to be carried out precisely and efficiently. The need is for algorithms with very high accuracy because it is an issue of life or death. However powerful these data mining techniques are it has to be used with great care in the biomedical applications. Within the issue of knowledge integrity assessment, two biggest challenges are: (1) How to develop efficient algorithms for comparing content of two knowledge versions (before and after). This challenge require development of efficient algorithms and data structures for evaluation of knowledge integrity in the data set; and (2) How to develop algorithms for evaluating the influence of particular data modifications on statistical importance of individual patterns that are collected with the help of common classes of data mining algorithm.

8. Conclusion

This paper will illustrate the of data mining healthcare information system and the function of the above medicals. This will also extract the information about the application user are used. This application is computer technical. Our future directions are to enhance the predictions using hybrid system. In future we will use for specific any techniques and try

to find out of efficiency and accuracy of result which we will examine on various healthcare system.

9. References

- [1] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2011.
- [2] C. T. Su, P. C. Wang, Y. C. Chen, and L. F. Chen, "Data mining techniques for assisting the diagnosis
- [3] C. T. Su, P. C. Wang, Y. C. Chen, and L. F. Chen, "Data mining techniques for assisting the diagnosis
- [4] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*, 2nd ed. Wiley-IEEE Press, 2011.
- [5] P. Berkhin, "A Survey of Clustering Data Mining," *Group. Multidimens. Data*, no. c, pp. 25–71, 2006.
- [6] T. P. Hong, K. Y. Lin, and S. L. Wang, "Fuzzy data mining for interesting generalized association rules," *Fuzzy Sets Syst.*, vol. 138, no. 2, pp. 255–269, 2003.
- [7] D. R. Hardoon, S. Sandor R., and S. John R., "Canonical Correlation Analysis: An Overview with Application to Learning Methods," *J. Neural Comput.*, vol. 16, no. 12, pp. 2639 – 2664, 2004.
- [8] J.Yanqing & H.Ying & J.Tran & P.Dews & A.Mansour & R.Michael Massanari, "MINING INFREQUENT CAUSAL ASSOCIATIONS IN ELECTRONIC HEALTH DATABASES", 11th IEEE International Conference on Data Mining Workshops, 2011.in associatred
- [9] Z. Wu and C. Li, "L0-Constrained Regression for Data Mining," pp. 981–988, 2007.
- [10] K.Sharmila & Dr.S.A.Vethamanickam, "SURVEY ON DATA MINING ALGORITHM AND ITS
- [11] Z. Wu and C. Li, "L0-Constrained Regression for Data Mining," pp. 981–988, 2007.