Indian J.Sci.Res. 17(2): 580-583, 2018 ISSN: 0976-2876 (Print)
ISSN: 2250-0138(Online)

DATA MINING TECHNIQUES IN HEALTHCARE

[1] P. Regina, [2] B. Kiranmayi, [3]P.Vandana
[1] Associate professor, Master of Computer Applications, Aurora PG College Uppal, Hyderabad
[2] Senior Assistant Professor, Master of Computer Applications, Aurora PG College Uppal,Hyderabad
[3] Assistant Professor, Master of Computer Applications, Aurora PG College Uppal, Hyderabad

***Abstract*:**Data mining applications can greatly assist all those involved in the healthcare industry. For example, data mining can help healthcare organizations make customer relationship management decisions, physicians to identify effective treatments and best practices, and patients to receive better and more affordable healthcare services and healthcare insurers to detect fraud and abuse. Data mining provides the methodology and technology to transform the heaps of data into useful information for decision making. The objective of this article is to explore relevant data mining applications by first examining data mining methodology and techniques; then, classifying potential data mining applications in healthcare major areas such as detection of diseases, the evaluation of treatment effectiveness, management of healthcare. It also gives the benefits of using Data Mining techniques in the computer-aided diagnosis focusing on the cancer detection, in order to help doctors to make optimal decisions quickly and accurately.

KeyWords: Data Mining, Health care, cancer detection and treatment

I. Introduction

Broadly speaking, the goals of data mining can be classified into two categories: description and prediction [1]. Descriptive data mining attempts to discover implicit and previously unknown knowledge, which can be used by humans in making decisions. In this case, data mining is part of a larger knowledge discovery process that includes data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and presentation of discovered knowledge to end-users. To arrive at usable results, it is essential that the discovered patterns are comprehensible by humans. Typical descriptive data mining tasks are unsupervised machine learning problems such as mining frequent patterns, finding interesting associations and correlations in data, cluster analysis, outlier analysis, and evolution analysis.

Predictive data mining seeks to find a model or function that predicts some crucial but yet unknown property of a given object, given a set of currently known properties. In prognostic data mining, for instance, one seeks to predict the occurrence of future medical events before they actually occur, based on patients' conditions, medical histories, and treatments [2]. Predictive data mining tasks are typically supervised machine learning problems such as regression and classification. Well-known supervised learning algorithms are decision trees, rule-based classifiers, Bayesian classifiers, linear and logistic regression analysis, artificial neural networks, and support vector machines.

Data mining is an emerging and interdisciplinary field, drawing from fields such as database systems, data warehouses, machine learning, statistics, signal analysis, data visualization, information retrieval, and high performance computing. It has been successfully applied in diverse areas such as marketing, finance, engineering, security, games, and science.

There is a tremendous opportunity for Data mining methods to potentially help all physicians who deal with the flood of patient information and scientific knowledge, by helping them to interpret complex diagnostic tests, by combining information from multiple sources (sample movies, images, clinical data, proteomics, scientific knowledge),by analyzing the common findings and patterns in the diagnosis to provide patient-specific prognosis.

A. Data Mining

In association, the objective is to determine which variables go together. For example, market-basket analysis (the most popular form of association analysis) refers to a technique that generates probabilistic statements such as, "If patients undergo treatment T, there is a 0.25 probability that they will exhibit symptom S." Such information can be useful for investigating associative relationships in healthcare. With clustering, the objective is to group objects, such as patients, in such a way that objects belonging to the same cluster are similar and objects belonging to different clusters are dissimilar.

B. Data mining in Health care

Health care data mining techniques are mostly predictive in nature and attempt to derive patterns that use patient-specific information to predict a patient's diagnosis, prognosis, or any other outcome of interest and to thereby support decision-making [3]. Historically, diagnostic applications have received more attention [4], but in the

[1]Corresponding Author

last decade prognostic models are becoming more popular [2,5].

Huge amount of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. Data mining improves decision-making by discovering patterns and trends in large amounts of complex data. Such analysis has become increasingly essential. Insights gained from data mining influences cost, revenue, and operating efficiency while maintaining a high level of care [6]. Healthcare organizations that perform data mining are better positioned to meet their long-term needs. Data can be a great asset to healthcare organizations, but they have to be first transformed into information, which is possible with Data Mining.

C. Data Mining Applications in Healthcare

There is great potential for data mining applications in healthcare. Generally, these can be grouped as the detection of disease, evaluation of treatment effectiveness; management of healthcare; customer relationship management;

a)Non-invasive diagnosis and decision support*:*

Some diagnostic and laboratory procedures are invasive, costly and painful to patients. An example of this is conducting a biopsy in women to detect cervical cancer. Thangavel et al (2006) used the K-means clustering algorithm to analyze cervical cancer patients and found that clustering found better predictive results than existing medical opinion. They found a set of interesting attributes that could be used by doctors as additional support on whether or not to recommend a biopsy for a patient suspected of having the cervical cancer.

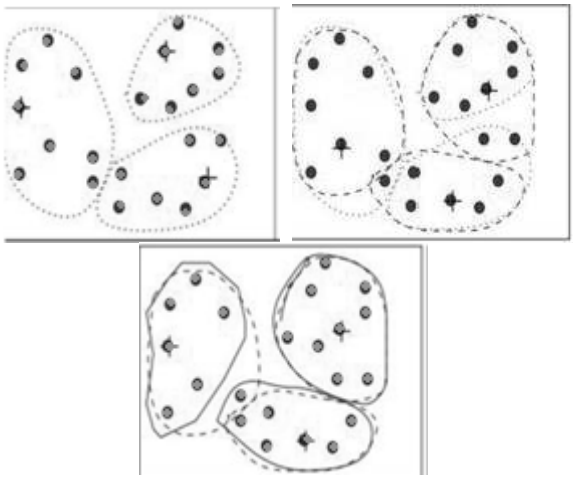

Figure 1: Clustering of a set of objects based on

K-means method

Recently developed methods of management in cancer diseases to replace palpation include a routine use of biopsy of the affected organ. However, biopsy is an invasive method, with inherent complications that may cause even the death of the patient. Consequently, the use of non-invasive alternatives is highly necessary

Gorunescu (2009) described how computer computer-aided diagnosis (CAD) and endoscopic ultrasonographic elastography (EUSE) were enhanced by data mining to create a new noninvasive cancer detection. In the traditional approach, doctors look at the ultrasound movie and decide on whether a patient is to be subjected to a biopsy.

Endoscopic ultrasonographic elastography (EUSE) is a recent elasticity imaging technique that reveals directly the physical properties of tissues. The method characterizes the difference of hardness between diseased tissue and normal tissue. This information can recently be obtained during real-time scanning. In the images resulted after scanning, colors express the difference of elasticity between healthy and diseased tissue [15].Different elasticity values are marked with different colors (on a scale of 1 to 255) and the EUSE information is shown as color movie. Technically, a EUSE sample movie (dynamic image) consists in a sequence of 125 frames (static images). The system uses by default a rainbow color coded map *red-green-blue* (RGB). Unfortunately, this methodology exhibits a major disadvantage due to the subjective means in which the human factor may analyze a large range of color nuances nuances, based on which an objective decision regarding the type of the tumor is to be taken. More than often, there are significant differences in the perception of close nuances, this fact resulting in individual decisions that are different from one doctor to another.

The physician's judgment is primarily subjective, depending mostly on the her interpretation of the ultrasound video. Gorunescu approached this problem in a different way, using data mining. He did not study patient demographics. Instead his team focused on the ultrasound movies. They first trained a classification algorithm using a multi-layer perceptron (MLP) on known cases of malignant and benign tumours.

The model analyzed the pixels and their RGB content to find sufficient patterns to distinguish between malignant and benign tumours. Then the team applied the resulting model to other cases. They found that their model resulted to high accuracy in diagnosis with only a small standard deviation.

Effectiveness of the Treatment. Data mining applications are developed to evaluate the effectiveness of medical treatments. By comparing and contrasting causes, symptoms, and courses of treatments, data mining can deliver an analysis of which courses of action prove effective. For example, the outcomes of patient groups

treated with different drug regimens for the same disease or condition can be compared to determine which treatments work best and are most cost-effective[7].

Other data mining applications related to treatments include associating the various side-effects of treatment, collating common symptoms to aid diagnosis, determining the most effective drug compounds for treating sub-populations that respond differently from the mainstream population to certain drugs, and determining proactive steps that can reduce the risk of affliction.

Data overload. There is a wealth of knowledge to be gained from computerized health records. Yet the overwhelming bulk of data stored in these databases makes it extremely difficult, if not impossible, for humans to sift through it and discover knowledge (Cheng, et al 2006).

In fact, some experts believe that medical breakthroughs have slowed down, attributing this to the prohibitive scale and complexity of present-day medical information. Computers and data mining are best suited for this purpose. (Shillabeer and Roddick 2007).

Prevention of hospital errors. When medical institutions apply data mining on their existing data, they discover new, useful and potentially life-saving knowledge that otherwise would have remained inert in their databases. For instance, an ongoing study on hospitals and safety found that about 87% of hospital deaths could have been prevented, had hospital staff (including doctors) been more careful in avoiding errors (HealthGrades Hospitals Study 2007). By mining hospital records, such safety issues could be flagged and addressed by hospital management and government regulators.

Early detection and/or prevention of diseases.. Cao et al (2008) described the use of data mining as a tool to aid in monitoring trends in the clinical trials of cancer vaccines. By using data mining and visualization, medical experts could find patterns and anomalies better than just looking at a set of tabulated data.

II. Data Mining Approach-Neural Network To Detect Cancer Tissues

In order to solve the EUSE deficiency mentioned above, regarding the way in which the images are interpreted and, finally, leading to major consequences as concerns the given diagnosis, we propose the employment of both the exploratory data analysis of the EUSE digitalized sample movies and the neural networks that will be trained to understand how to classify tumors as benign or malignant, based on the analysis of previously digitalized images.

Firstly, in order to apply the neural network methodology to differentiate between the sample movies, characterizing either benign or malignant tumors, we firstly need to digitalize them.

Since the corresponding EUSE sample movie (dynamic image) consists in a sequence of 125 frames (static images) displaying 255 colors, then, using the public domain Java based image processing tool, a number of 125 (hue) histograms are obtained, providing the distributions of (hues) colors in each frame. Thus, from mathematical point of view, to each patient corresponds a 125×255 matrix (*aij*), each row representing a certain frame of the sample movie and each column representing a pixel color.

Secondly, departing from the corresponding database that contains the digitalized forms of the scanned images of different tumoural tissues, displayed as sample movies, together with the corresponding diagnosis that was established without any doubt by doctors, the artificial neural networks are trained to learn to associate a certain color pattern to the corresponding diagnosis (benign/malignant). The power of this novel methodology of detecting the cancer in a noninvasive way comes to life when the digitalized sample movie of a new (non-diagnosed) patient is presented. In this case, the neural network gives the output that corresponds to a taught pattern that is least different from the given pattern. Using this novel neural network approach, the physicians will combine the opportunity given by the neural networks approach and their expertise to successfully predict the malignancy of a given tumor.

Databases usually include a query facility, and the database community has a tendency to view data mining methods as more complicated types of database queries. For example, standard query tools can answer questions such as, “How many surgeries resulted in hospital stays longer than 10 days?” Data mining is valuable for more complicated queries such as, “What are the important preoperative predictors of excessive length of stay?” Data mining techniques can be implemented retrospectively on massive data in an automated matter, whereas traditional statistical methods used in epidemiology require custom work by experts. Traditional methods generally require a certain number of predefined variables, whereas data mining can include new variables and accommodate a greater number of variables.

Traditional methods, such as statistical process control based on various underlying probability distribution functions, have been successfully implemented in hospital infection control.1-8 Data mining techniques have been implemented separately, and some of these are described below. Direct comparison of traditional statistical methods with data mining would require competitive results on the same data. Application of either statistical or data mining techniques requires substantial human effort, and collaboration, rather than competition, needs to occur between the two fields. As more statisticians become involved in data mining, the two fields could contribute to each other more effectively by building on each other’s

strengths to create synergy than by having a "bake off" or taking an antagonistic approach.

Many database vendors are moving away from providing stand-alone data mining workbenches toward embedding the mining algorithms directly in the database. This process is known as "in place data mining" and it enables more efficient data management and processing.

III. Conclusion

This survey of data mining applications in medicine and health care provided only an overview of current practices. Data mining applications in healthcare can have tremendous potential and usefulness. However, the success of healthcare data mining hinges on the availability of clean healthcare data. In this respect, it is critical that the healthcare industry consider how data can be better captured, stored, prepared, and mined. The standardization of clinical vocabulary and the sharing of data across organizations is required, to enhance the benefits of healthcare data mining applications.

Refrences

[1]. Hand DJ, Mannila H, Smyth P. Principles of Data Mining. Cambridge, Massachusetts: MIT Press; 2001.

[2]. Abu-Hanna A, Lucas PJ. Prognostic models in medicine: AI and statistical approaches. Methods Inf Med 2001; 40 (1): 1–5.

[3]. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform 2008; 77 (2): 81–97.

[4]. Lavrac N, Kononenko I, Keravnou E, Kukar M, Zupan B. Intelligent data analysis for medical diagnosis: using machine learning and temporal abstraction. AI Commun 1998; 11: 191–218.

[5]. Pfaff M , Weller K, Woetzel D, Guthke R, Schroeder K, Stein G, Pohlmeier R, Vienken J. Prediction of cardiovascular risk in hemodialysis patients by data mining. Methods Inf Med 2004; 43 (1): 106–113.

[6]. Silver, M. Sakata, T. Su, H.C. Herman, C. Dolins, S.B. & O'Shea, M.J. (2001). Case study: how to apply data mining techniques in a healthcare data warehouse. Journal of Healthcare Information Management, 15(2), 155-164.

[7]. Kincade, K. (1998). Data mining: digging for healthcare gold. Insurance & Technology, 23(2), IM2-IM7.