

HEART ATTACK PREDICTION USING DATA MINING TECHNIQUE

Dr. G. Rasitha Banu MCA., M.Phil., Ph.D.¹, J.H. BOUSAL JAMALA MCA.²

¹Assistant Professor, Department of Health Information Management and Technology, Faculty of Public Health and Tropical Medicine, Jazan University, Kingdom of Saudi Arabia.

²MPhil, Department of Computer Science, Mother Teresa Women's University Chennai, India

Abstract-- Data mining techniques are used to analyze this rich collection of data from different perspectives and deriving useful information. This project intends to design and develop diagnosis and prediction system for heart diseases based on predictive mining. Heart disease is a term that assigns to a large number of medical conditions related to heart. These medical conditions describe the abnormal health conditions that directly influence the heart and all its parts. Heart disease is a major health problem in to days time. This paper aims at analyzing the various data mining techniques introduced in recent years for heart disease prediction. Cardiovascular disease remains the biggest cause of deaths worldwide. This paper, a new unsupervised classification system is adopted for heart attack prediction at the early stage using the patient's medical record. The information in the patient record are preprocessed initially using data mining techniques and then the attributes are classified using a Fuzzy C means classifier. In the classification stage 13 attributes are given as input to the Fuzzy C Means (FCM) classifier to determine the risk of heart attack. FCM is an unsupervised clustering algorithm, which allows one piece of data to belong to two or more clusters.

Keywords-- FUZZY C algorithm, cardiovascular disease, Clustering Classification, Predictive data mining.

I. INTRODUCTION

Medical Informatics is the applied sciences of medicine and information technology, which provides measurable improvements in both quality of care and effectiveness. Information technologies are playing a crucial role in advancing the science of quality measurement but more can be done to apply it to quality improvement. The Health care provides various services which are used to: (1) improve quality and efficiency; (2) engage patients and families; improve care coordination, and population and public health; and (3) Maintain privacy and security of patient health information. The most predominant health issue is heart failure which occurs especially in old patients because of diet, non-steroidal anti-inflammatory drugs and will leads even towards death. One of the commonly occurred heart diseases is Cardio vascular disease. Thus it is highly essential to predict such diseases through suitable symptoms. There are various types of algorithms which are present for the prediction of heart diseases. Cardiovascular disease is a kind of serious health imperiling and frequent happening disease. The world health organization has estimated that 12 million deaths occur worldwide, every year due to the cardiovascular disease. Advances in the field of medicine over the past few decades enabled the identification of risk factors that may contribute toward the cardiovascular diseases. The most common cause of heart disease is narrowing or blockage of the coronary arteries, the blood vessels that supply blood to the heart itself. This is called coronary artery disease and happens slowly over time. It's the major reason people have heart attacks. A blockage that is not treated within a few hours causes the affected heart muscle to die. During about 30 percent of all heart attacks, the patient

experiences no symptoms. However, unmistakable signs of the attack remain in the bloodstream for days. Medical diagnosis is an important but complicated task that should be performed accurately and efficiently and its automation would be very useful. All doctors are unfortunately not equally skilled in every sub specialty and they are in many places a scarce resource. A system for automated medical diagnosis would enhance medical care and reduce costs. With so many factors to analyze for a diagnosis of heart attacks, physicians generally make a diagnosis by evaluating a patient's current test results. Previous diagnoses made on other patients with the same results are also examined by physicians. These complex procedures are not easy. Therefore, a physician must be experienced and highly skilled to diagnose heart attacks in a patient. Thus the effort to utilize knowledge and experience of numerous specialists and clinical screening data of patients collected in databases to facilitate the diagnosis process is considered a valuable system that is the integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome.

II. METHODOLOGY

The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System with Fuzzy C Means Clustering algorithm using historical heart disease databases to make intelligent clinical decisions which traditional decision support systems. Fuzzy c-means clustering algorithms for predicting the heart attack diseases. In these fuzzy clustering algorithms, the membership degree is associated to the values of the features in the clusters for the cluster centers instead of being associated to the patterns in each cluster.

III. RESEARCH FINDINGS

DATA MINING ALGORITHMS AND TECHNIQUES

DATA MINING

DATA MINING in health care has become increasingly popular because it can improve our patient care by early detecting of disease supports helping care providers for treatment programs and reduces the cost of health care.

FUZZY C

To overcome this problem using Fuzzy C Means Clustering algorithm. For finding the risk of heart attack of a patient using the profiles collected from the patients. With the proper adaptation of FCM classifies, the method can thus evolve an optimum number of clusters and finds the abnormal and normal cases efficiently. Initially pre-processing of the data is done to remove all the duplicate records and add missing data. In the classification stage, a FCM classifier is used to classify the data as heart disease present or not.

Modules:

Records acquirement:

In this module, heart attack dataset is obtained from UCI centre for machine knowledge and intelligent system. The data have been collected from 270 patients. The data base contains 73 attributes but we use 13 attributes. That attributes are age, sex, chest pain, Rest BP, cholesterol, sugar, ECG, Max heart range, Angina, Old peak, Sis lope, vessels, thal.

Pre-processing:

Data pre-processing is an important step in the data mining process. Data-collection methods are often insecurely controlled, resulting in out-of-range values, impractical data combinations (e.g., Sex: Male, Pregnant: Yes), missing values, etc. In this method, data's are extracted from UCI centre dataset And this module remove to all duplicate records and missing records.

Fuzzy C Means (FCM) classifier

In fuzzy clustering, every point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster, may be *in the cluster* to a lesser degree than points in the center of cluster. An overview and comparison of different fuzzy clustering algorithms is available.^[1]

Any point x has a set of coefficients giving the degree of being in the k th cluster $w_k(x)$. With fuzzy c -means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster:

$$c_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}.$$

The degree of belonging, $w_k(x)$, is related inversely to the distance from x to the cluster center as calculated on the previous pass. It also depends on a parameter m that controls how much weight is given to the closest center. The fuzzy c -means algorithm is very similar to the k -means algorithm

Choose a number of clusters.

- Assign randomly to each point coefficients for being in the clusters.
- Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than, the given sensitivity threshold) :
 - Compute the centroid for each cluster, using the formula above.
 - For each point, compute its coefficients of being in the clusters, using the formula above.

The algorithm minimizes intra-cluster variance as well, but has the same problems as k -means; the minimum is a local minimum, and the results depend on the initial choice of weights.

Using a mixture of Gaussians along with the expectation-maximization algorithm is a more statistically formalized method which includes some of these ideas: partial membership in classes.

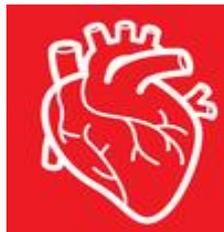
CLUSTERING

Clustering can be said as identification of similar classes of objects. By using clustering techniques, we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification.

CLASSIFICATION

In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. For a fraud detection application, this would include complete records of both fraudulent and valid activities determined on a record-by-record basis.

Heart disease



The heart

Your heart is the size of your fist and the strongest muscle in your body. Your heart started beating about three weeks after you were conceived. If you live to be 70 years

heart will have beat two and a half billion times. Each heartbeat pumps blood around the body, pushing it from the left heart chambers, through arteries of ever-decreasing size, finally reaching the capillaries in all parts of the body.

Once your body has taken oxygen and nutrients from the blood, it is returned to the heart via the veins to the right chambers of the heart. On its way back, the blood passes through the liver and waste products are removed.

As marvelous as this system is, it is very vulnerable to damage from the things we do to it, like smoking, eating an unhealthy diet or putting it under stress. Or you may be born with a heart condition. When your heart's functions become compromised, this is known as cardiovascular disease, a broad term that covers any disorder to the system that has the heart at its centre.

Causes and impact of heart diseases

According to WHO report Global atlas on cardiovascular disease prevention and control states that cardiovascular disease (CVDs) are the leading causes of death and disability in the world. Although a large proportion of CVDs is preventable, they continue to rise mainly because preventive measures are inadequate. Over 17.3 million An estimated 17.3 million people died from CVDs in 2008, Over 80% of CVD deaths take place in low- and middle-income countries, 23.6 million By 2030, almost 23.6 million people will die from CVDs.

Protect heart health

- Tobacco use, an unhealthy diet, and physical inactivity increase the risk of heart attacks and strokes.
- Engaging in physical activity for at least 30 minutes every day of the week will help to prevent heart attacks and strokes.
- Eating at least five servings of fruit and vegetables a day, and limiting your salt intake to less than one teaspoon a day, also helps to prevent heart attacks and strokes.

Cardiovascular Diseases

Cardiovascular disease is caused by disorders of the heart and blood vessels, and includes coronary heart disease (heart attacks), cerebrovascular disease (stroke), raised blood pressure (hypertension), peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure. The major causes of cardiovascular disease are tobacco use, physical inactivity, an unhealthy diet and harmful use of alcohol. These are the three causes of heart diseases (1) chest pain (2) stroke and (3) heart attack.[14] To prevent and identification of these diseases different techniques of data mining is used through this easily find out heart related diseases and this is the aim of this research studies. Heart disease is the leading cause of death all over the world in the past ten years. Several researchers are using statistical and data mining tools to help health care professionals in the diagnosis of heart disease.

IV. CONCLUSION

In this paper, we have proposed an FCM clustering algorithm for finding the risk of heart attack of a patient using the profiles collected from the patients. With the proper adaptation of FCM classifies, the method can thus evolve an optimum number of clusters and finds the abnormal and normal cases efficiently. Initially pre-processing of the data is done to remove all the duplicate records and add missing data. In the classification stage, a FCM classifier is used to classify the data as heart disease present or not. The results of classification experiment, performed over a data sets obtained from 270 patients, shows that the classifier has achieved better accuracy than most of the existing algorithms. The performance of the proposed FCM is proved to be a well known approach in terms of accuracy.

REFERENCE

- [1] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques" IEEE Conference, 2008, pp 108-115.
- [2] Carolas Ordonez "Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction" IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 2, April 2006.
- [3] Shanthakumar, Patil, Y.S., Kumaraswamy "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network". European Journal of Scientific Research Vol. 31, No. 04, 2009, 642-656 .
- [4] Palaniappan, S., Awang, R., 2008. Intelligent heart disease prediction system using data mining techniques. International Journal of Computer Science and Network Security 8 (8), 108–115.
- [5] Shanthakumar B. Patil, Y.S., Kumaraswamy "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network". European Journal of Scientific Research Vol. 31, No. 04, 2009, 642-656.
- [6] Yang, M.-S., Hwang, P.-Y., Chen, D.-H., 2004. Fuzzy clustering algorithms for mixed feature variables. Fuzzy Sets Systems 141, 301 – 317.
- [7] Ross, Timothy J., 2004. Fuzzy Logic with Engineering Applications, John Wiley & Sons, Second Edition.
- [8] Lanhai L. Comparison of conventional and fuzzy land classification and evaluation techniques in Oxford shire England. Int Agric Eng J 1998;7:1
- [9] Asha Rajkumar, G. Sophia Reena, Diagnosis Of Heart Disease Using Data Mining Algorithm, Global Journal of Computer Science and Technology 38 Vol. 10 Issue 10 Ver. 1.0 September 2010.

