ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)
Vol. 2, Issue 3 (July - Sept. 2014)
International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014)

Domain Analysis Over Cardiac Disease by Using Various Techniques of Data Mining

Sen Umesh Kumar
Guest Faculty, Dept.of Comp.Sc. & Application., Govt. Mgm PG College, Itarsi, M.P., India

Abstract
Healthcare industry having the large amount of data, however unfortunately most of them data is is not mined to find out hidden information in form of data. Well, data mining techniques can be used to discover hidden facts in their datas , which is related to cardiac disease. This research paper would be helpful to medical practitioner to make effective decision.A recent survey says about 25 % of death in the age group of 25-69 years occur because of cardiac disease. Disease diagnosis is one of the application , where data mining techniques are providing successful result. Well, this research paper proposed to find out the cardiac disease prediction through various data mining techniques namely , support vector machine(SVM) , Genetic algorithm , rough set theory, decision tree and neural network.

Keywords
Data mining, Heart disease, SVM, rough set techniques and decision tree.

I. Introduction

Data mining is the one of the core step, which results in the discovery of hidden but useful knowledge from massive databases. Data mining is the non trivial extraction of implicit previously unknown and potentially useful information about data. The development of Information Technology has generated large amount of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis. The main objective of our project is to develop a prototype Intelligent Heart Disease Prediction System (IHDPS) using three data mining modelling techniques, namely, Decision Trees, Naïve Bayes and Neural Network. IHDPS can discover and extract hidden knowledge (patterns an relationships) associated with heart disease from a historical heart disease database. A major challenge facing healthcare organizations like as hospitals, medical centers is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable.

II. Cardiac Disease

As we know that Life is completely dependent on efficient working of the heart. The term Heart disease refers to disease of heart & blood vessel system within it.
There are number of factors which increase the risk of Heart disease :

- Family History Of Heart Disease
- Smoking
- Cholesterol
- Poor Diet
- High Blood Pressure
- High Blood Cholesterol
- Obesity
- Physical Inactivity
- Hyper Tension.

Nowadays, in the world Heart disease is the major cause of deaths. The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to the Heart diseases. In 2008, 17.3 million people died due to Heart Disease. Over 80% of deaths in world are because of Heart disease. WHO estimated by 2030, almost 23.6 million people will die due to Heart disease.
Symptoms of a Heart Attack:
Symptoms of a heart attack can include:

- Discomfort, pressure, heaviness, or pain in the chest, arm, or below the breastbone.
- Discomfort radiating to the back, jaw, throat, or arm.
- Fullness, indigestion, or choking feeling (may feel like heartburn).
- Sweating, nausea, vomiting, or dizziness.
- Extreme weakness, anxiety, or shortness of breath.
- Rapid or irregular heartbeats.

1. What to do while Cardiac Attack:

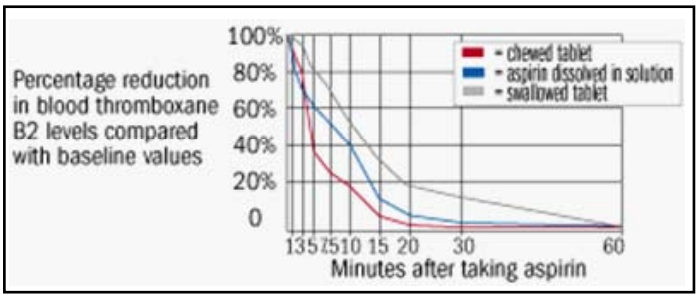
A. CPR
You use cardiopulmonary resuscitation (CPR) to revive someone who has stopped breathing or whose heartbeat has stopped (cardiac arrest). Not everyone who has a heart attack needs CPR because not all heart attacks cause the heart to stop beating.If someone suddenly collapses or passes out and is not responding to you,

B. Immediately call emergency medical services (EMS) as 108 or whatever.

C. If there is available life saving drug as Aspirin provided immediately that can help prevent heart attacks in patients with coronary artery disease and in healthy men over 50 years of age. Only low doses, between 81 and 325 mg a day, are needed. But people who think they may be having an attack need an extra 325 mg of aspirin, and they need it as quickly as possible. For the best results, chew a single full-sized 325-mg tablet, but don't use an enteric-coated tablet, which will act slowly even if chewed. And don't forget to call emergency medical services(EMS) as 108 or whatever, then your doctor. It's a contemporary update on the old reminder to take two aspirin and call in the morning — and it's good advice to chew over.

Antiplatelet effect of chewed, swallowed, and dissolved aspirin

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)



Chewing aspirin hastens its antiplatelet effect, as measured by the reduction in blood thromboxane B2 levels. It took only 5 minutes for patients who chewed aspirin to achieve a 50% reduction in baseline levels, versus almost 8 minutes after they took it in a solution and 12 minutes after they swallowed it whole.

Source: American Journal of Cardiology Vol. 84, p. 404.

III. Literature Review

Numerous studies have been done that have focus on diagnosis of heart disease. They have applied different data mining techniques for diagnosis & achieved different probabilities for different methods.

- An Intelligent Heart Disease Prediction System (IHDPS) is developed by using data mining techniques Naive Bayes, Neural Network, and Decision Trees was proposed by SellappanPalaniappan. Each method has its own strength to get appropriate results. To build this system hidden patterns and relationship between them is used. It is web based, user friendly & expandable.
- To develop the multi-parametric feature with linear and nonlinear characteristics of HRV (Heart Rate Variability) a novel technique was proposed by HeonGyu Lee. To achieve this, they have used several classifiers e.g. Bayesian Classifiers, CMAR (Classification based on Multiple Association Rules), C4.5 (Decision Tree) and SVM (Support Vector Machine).
- Researchers have been applying various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., to help health care professionals with improved accuracy in the diagnosis of heart disease.

IV. Data Mining Techniques Over Cardiac Disease: A Prediction

The five different data mining classification techniques, i.e. Neural Networks, Decision Trees, Support Vector Machine(SVM), Dataset technique and Naive Bayes theorem's are used to analyze the dataset.

A. Neural Networks in Data Mining

Artificial neural networks (ANNs) are commonly known as biologically inspired, highly sophisticated analytical techniques, capable of modeling extremely complex non-linear functions. One of popular ANN architecture is called multi-layer perceptron (MLP) with back-propagation (a supervised learning algorithm). The MLP is known to be a powerful function approximator for prediction and classification problems. The MLP is essentially the collection of nonlinear neurons (perceptrons) organized and connected to each other in a feedforward multi-layer structure. The primary task of neurons in input layer is the division of input signal xi among neurons in hidden layer. Every neuron in hidden layer adds up its input signals xi once it weights them with the strength of the respective connections wji from the input layer and determines its output yj as a function f of the sum, given as

$$Yj = f(\Sigma Wji\ Xi)$$

1. Supervised Learning

It is a simple model, in which the networks computea response to each input and then compare it with target value. If the computed response differs from target value, the weights of the network are adapted according to a learning rule.
e.g.: (A) Single-layer perceptron, (B) Multi-layer perceptron.

2. Unsupervised Learning

These networks learn by identifying special features in the problems they are exposed to. e.g.: Self-organizing feature maps.
In medical field, decision making is done by neural network because they provide more accurate results.

B. Decision Tree in Data Mining

J48 algorithm uses pruning method to build a tree. Pruning is a technique that reduces size of tree by removing over fitting data, which leads to poor accuracy in predications. The J48 algorithm recursively classifies data until it has been categorized as perfectly as possible. This technique gives maximum accuracy on training data. The overall concept is to build a tree that provides balance of flexibility & accuracy.Decision trees are powerful classification algorithms. Popular decision tree algorithms include Quinlan's ID3, C4.5, C5, and Breiman et al.'s CART.

C. Support Vector Machine (SVM)

The SVM is a state-of-the-art maximum margin classification algorithm rooted in statistical learning theory. SVM is method for classification of both linear and non-linear data. It uses a non-linear mapping to transform the original training data into a higher dimension. Within this new dimension it searches for linear optimal separating hyperplane. With an appropriate non-linear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane. The SVM find this hyperplane using support vectors and margins . SVM performs classification tasks by maximizing the margin separating both classes while minimizing the classification errors.

V. Dataset Technique in Data Mining:

To compare these data mining classification techniques Cleveland cardiovascular disease dataset from UCI repository was used. The dataset has 13 attributes , which is described in table 1 below :

Table 1: Attributes of Cardiovascular disease dataset:

S . NO.	Attributes	Description	Values
1.	Age	Age in years	Continous
2.	Sex	Male or female	1=male 0=female
3.	CP	Chest pain type	1=typical type 2=typical type angina 3=non-angina pain 4=asymptomatic

4.	Thestbps	Resting blood pressure	Continuous value in mm hg
5.	Chol	Serum cholesterol	Continuous value in mm/dl
7.	Fbs	Fasting blood sugar	1 ≥120 mg/dl 0 ≤120 mg/dl
8.	Thalach	Maximum heart rate achieved	Continuous value
9.	Exang	Exercise induced angina	0= no 1 = yes
10.	Oldpeak	ST depression induced by exercise relative to rest	Continuous value
11.	Slope	Slope of the peak exercise ST Segment	1 = unsloping 2 = flat 3 = downsloping
12.	Ca	Number of major vessels colored by Floursopy	0-3 value
13.	Thal	Defect type	3 = normal 6 = fixed 7 = reversible defect

All the research papers referred above have used 13input attributes for prediction of Heart disease. For getting more accurate results 2 more parameters are used i.e. obesity and smoking.

VI. Naive Bayes Theorem in Data Mining

Naive Bayes classifier is based on Bayes theorem. This classifier algorithm uses conditional independence, means it assumes that an attribute value on a given class is independent of the values of other attributes.

The Bayes theorem is as follows:

Let X={x1, x2, ...,xn} be a set of n attributes.

In Bayesian, X is considered as evidence and H be some hypothesis means, the data of X belongs to speci fic class C.

We have to determine P (H|X), the probability that the hypothesis H holds given evidence i.e. data sample X.

According to Bayes theorem the P (H|X) is expressed as

P (H|X) = P (X| H) P (H) / P

VII. Conclusion

Hence the objective is to study the various data mining techniques available to predict the heart disease and to compare them to find out the accuracy over prediction. There are different data mining techniques that can be used for the identification and prevention of cardiovascular disease. In this research paper, we have presented Heart disease prediction system (HDPS) using data mining and artificial neural network (ANN) techniques. The system extracts hidden knowledge from a historical heart disease database.

References

[1]. Preventing Chronic Disease: A Vital Investment. World Health Organization Global Report, 2005.

[2]. Global Burden of Disease. 2004 update (2008). World Health Organization.

[3]. Yanwei, X.; Wang, J.; Zhao, Z.; Gao, Y., "Combination data mining models with new medical data to predict outcome of coronary heart disease".

[4]. Srinivas, K.," Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques".

[5]. Frawley and G. Piatetsky -Shapiro, Knowledge Discovery in Databases

[6]. SellappanPalaniappan, RafiahAwang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques".

[7]. ShantakumarB.Patil, Y.S.Kumaraswamy "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network".

[8]. Fayyad, U: "Data Mining and Knowledge Discovery in Databases: Implications fro scientific databases".

[9]. Kaur, H., Wasan, S. K.: "Empirical Study on Applications of Data Mining Techniques in Healthcare".

[10]. Obenshain, M.K: "Application of Data Mining Techniques to Healthcare Data",

Profile

Umesh Kumar Sen, Education: MCA, Occupation: Working as a Guest Faculty in dept.of Computer Science & Application in Govt.MGM PG College, Itarsi(M.P.) 461111. Address: C/o Dept.of Computer Science & Application, Govt. MGM PG College, Nyas Colony, Itarsi(M.P.) 461111. Contact Details: 09479359248, -mail id: umeshkumarsen10@gmail.com