IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 5 Issue 4, April 2018
ISSN (Online) 2348 – 7968
www.ijiset.com

An Enhanced K-Means Algorithm for Classification of Cardiac Disease

R.Harish Kumar [1] DR.E.Venkatesan
PG Student, ,Bharath Institute of Higher Education and Research, Chennai-73,
India[1] harikumar.hk596@gmail.com
Assistant professor, Bharath Institute of Higher Education and Research, Chennai-73,
India[2]venkatelumalai12@yahoo.co.in

ABSTRACT:

Medical science industry has huge amount of data, but unfortunately most of this data is not mined to find out hidden information in data. Advanced data mining techniques can be used to discover hidden pattern in data. Models developed from these techniques will be useful for medical practitioners to take effective decision. In this research work, we have analyzed the performance of the classification rule algorithms namely PART based on K-Means Clustering algorithms. The k-means is the simplest, most commonly and good behavior clustering algorithm used in many applications. Firstly the preprocessed heart disease dataset is grouped using the K-means algorithm with the K =2 values on classes to cluster evaluation testing mode. After that data mining classification rule algorithms namely Projective Decision tree are analyzed on clustered relevant dataset. In our studies 10-fold cross validation method was used to measure the unbiased estimate of the prediction model. Accuracy of K-Means Clustering, PART and PART based on K-Means Clustering are 81.08%, 79.05% and 99.02% respectively.

INT RODUCT ION

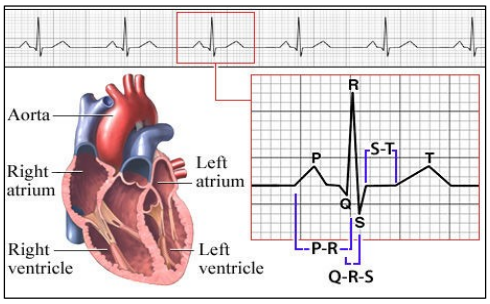
The most important aim of this paper is to study the different data mining techniques used in prediction of heart disease by using different data mining tools. Life is dependent on competent functioning of heart, because heart is necessary part of our body. If function of heart is not suitable, it will affect the other body parts of human such as brain, kidney etc. Heart disease is a disease that affects on the function of heart. There are number of factors which increases risk of heart disease. At the present days, in the world heart disease is the main cause of deaths. The World Health Organization (WHO) has expected that 12 million deaths occur

worldwide, every year due to the heart diseases. Prediction by using data mining techniques gives us accurate result of disease. HDPS (heart disease prediction system) can find out and extract hidden knowledge related with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus help healthcare analysts and practitioners to make intelligent clinical decisions which conventional decision support systems cannot. A few kinds of heart disease are cardiovascular diseases, heart attack, coronary heart disease and Stroke. Stroke is a type of heart disease; it is caused by narrowing, blocking, or hardening of the blood vessels that go to the brain or by high blood pressure. In previous studies of some researcher expressed their attempt on finding best approach for risk prediction model and here found best model by comparing those researcher's findings as survey. In this exclusive approach because different techniques scheduled and expressed in table bar to recognize accuracy rank of every. These techniques are selected based on their efficiency in the literature. This survey helps to understand the recent techniques concerned in risk prediction of heart disease at classification in data mining.

OVERVIEW:

A natural electrical system causes the heart muscle to contract and pump blood through the heart to the lungs and the rest of the body. The electrical potential generated by electrical activity in the cardiac tissue calls biosignal. An ECG translates the heart electrical activity into line tracings on paper which shown in Figure 1.1.The spikes and dips in the line tracings are called waves which are P wave, QRS complex, T wave and ST segment. pump blood through the heart to the lungs and the rest of the body as carried out by Melnyk and Silberman (2004), Gandelman (2006) and Otto et al. (2006). The P wave is a record of the electrical activity through the upper heart chambers (atria). The QRS complex is a record of the movement of electrical impulses through the lower heart chambers (ventricles). The ST segment corresponds to the time when the ventricle is contracting but no electricity is flowing through it. The ST segment usually appears as a straight, level line between the QRS complex and the T wave. Then, the T wave corresponds to the period when the lower heart chambers are relaxing electrically and preparing for their next muscle contraction.

Diagram of the Human Heart and an Example of Normal ECG Trace



The heart is a muscular pump made up of four chambers as shown in Figure 1.1. The two upper chambers are called atria, and the two lower chambers are called ventricles. A natural electrical system causes the heart muscle to contract and pump blood through the heart to the lungs and the rest of the body as carried out by Melnyk and Silberman (2004), Gandelman (2006) and Otto et al. (2006). The P wave is a record of the electrical activity through the upper heart chambers (atria). The QRS complex is a record of the movement of electrical impulses through the lower heart chambers (ventricles). The ST segment corresponds to the time when the ventricle is contracting but no electricity is flowing through it. The ST segment usually appears as a straight, level line between the QRS complex and the T wave. Then, the T wave corresponds to the period when the lower heart chambers are relaxing electrically and preparing for their next muscle contraction.

In the medical test using ECG, the heart disease detection is based on the difference wave signals that appear on the screen during the ECG test. The detection of the pulse is usually detected on the basis of the largest in a signal of PQRST of ECG signal.

The normal heart beat in a regular rhythm will show the line tracing of the PQRS and T wave looks normal. If there any obvious changes of the PQRST line tracing, it shows that the heart may having a problems. Comparison of overall ECG waveform pattern and shape enables doctors to identify diseases. The ECG remains the simplest noninvasive diagnostic method for various heart diseases.

In order to accurately characterize heart rate analysis, so a precise and reliable ECG waveform recognition procedure is necessary. The information and measurement waveform of the ECG signal is to extract features (characteristic) from the signal. The features are sufficiently representative of the physical process and the heart disease problem. Previous studies prove that frequency analysis sufficiently represents ECG waveforms; therefore frequency analysis is utilized to extract features from

the ECG signal. The shapes of the ECG waveforms of different persons are different, so the differences of the waveform can be used to identify the different individual's characteristic. The ECG signal can varies from person to person due to the differences in position, size anatomy of the heart, age, sex relative body weight and also chest configuration. In the past, many works of using algorithms for ECG analysis have been investigated and the successful results are achieved. However, mostly the works are regarding to the specific disease checking system with so many limitations. Nowadays, many researches of the ECG waveforms detection methods such as K-Mean, Artificial Neural Networks, Genetic Algorithms, Pattern Comparison, Wavelet Transform and other methods.

RELATED WORKS

We now compare our solution to related real-time episte-mologies methods. A litany of prior work supports our use of the evaluation of compilers [9]. We had our solution in mind before Davis published the recent well-known work on peer-to-peer symmetries. Sato and Davis [15], [2] suggested a scheme for architecting secure technology, but did not fully realize the implications of Boolean logic at the time [3,12,12,13,17]. This is arguably ill-conceived. Our method to cacheable epistemologies differs from that of Li and Watanabe [15] as well.

A major source of our inspiration is early work by William Kahan et al. on flexible technology. We believe there is room for both schools of thought within the field of independently saturated disjoint relational e-voting technology. Further, the original solution to this problem by J.H. Wilkinson was well-received; unfortunately, it did not completely overcome this obstacle. This work follows a long line of prior frameworks, all of which have failed [16]. Recent work by K. P. Martinez [11] suggests an algorithm for exploring consistent hashing, but does not offer an implementation. We plan to adopt many of the ideas from this existing work in future versions of LeyDirk.

METHODS AND METHODOLOGY

Introduction to Java:

Java is two things: a programming language and a platform.

The Java Programming Language

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 5 Issue 4, April 2018
ISSN (Online) 2348 – 7968
www.ijiset.com

Java is a high-level programming language that is all of the following:

- Simple
- Object-oriented
- Distributed
- Interpreted
- Robust
- Secure
- Architecture-neutral
- Portable
- High-performance
- Multithreaded
- Dynamic

Java is two things: a programming language and a platform.

-

The Java Programming Language

Java is a high-level programming language that is all of the following:

- Simple
- Object-oriented
- Distributed
- Interpreted
- Robust
- Secure
- Architecture-neutral
- Portable
- High-performance
- Multithreaded
- Dynamic

The relationship between our methodology

LeyDirk is composed of a virtual machine monitor, a hand-optimized compiler, and a virtual machine monitor. The client-side library contains about 9598 instructions of ML [4]. Our methodology is composed of a server daemon, a server daemon, and a virtual machine monitor. This is essential to the success of our work. Next, our methodology is composed of a collection of shell scripts, a collection of shell scripts, and a virtual machine monitor. One cannot imagine other methods to the implementation that would have made optimizing it much simpler.

We now discuss our performance analysis. Our overall evaluation seeks to prove three hypotheses:\

(1) that Boolean
logic no longer adjusts system design;

(2) that simulated annealing no longer adjusts system design; and finally

(3) that spreadsheets no longer influence complexity. An astute reader would now infer that for obvious reasons, we have intentionally neglected to refine energy. We hope to make clear that our monitoring the median interrupt rate of our mesh network is the key to our evaluation approach.

Future Work

In our future work, we have planned to design and develop an efficient heart attack prediction system with Patient Prescription Support using the web mining and data warehouse techniques. New algorithms and techniques are to be developed which overcome the drawbacks of the existing system. In future some privacy preserving technique can be induced for the rule generation in the classification technique. We intend to improve performance of these basic classification techniques by creating Meta model which will be used to predict cardiovascular disease in patient.

CONCLUSION:

Around 18 million people 7% of the Indians are affected by heart disease. Heart disease is mostly affected the person under the age of 65. In this paper, we have compared PART and PART based on K-Means Clustering algorithms which are very suitable for generating rules using decision tree classification technique. The classification rule generation algorithms generates classification rules which is both sensitive and non sensitive. There are different data mining techniques that can be used for the identification and prevention of cardiovascular disease among patients. Our studies showed that Part based on K-Means Clustering turned out to be best classifier for cardiovascular disease prediction

REERENCE

[1] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, Wei Din, ”Data Mining With Big Data”,IEEE Transactions on Knowledge and Data Engineering, Vol. 26,No. 1, January 2014

[2] Ankita Dewan, Meghna Sharma, ”Prediction of Heart Dis- eases Using A Hybrid Technique In Data Mining

Classifiscation", 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom).Jack Galilee, Dr. Ying Zhou, "A Study on Implementing Iterative Algorithms Using Big Data Frameworks", University of Sydney, School of Information Technologies,Faculty of Engineering and Information technologies,2014.

[3] Wullianallur Raghupathi, Viju Raghupathi, "Big data analytics in healthcare: promise and potential" , Raghupathi and Raghupathi, Health Information Science and Systems, 2:3, 2014.

[4] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications (0975,a 8887), Volume 17, a No.8, March 2011.

[5] XindongWu, Vipin Kumar,J.Ross, Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J, McLachlan, Angus Ng, Bing Liu, Philip S.Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, Dan Steinberg, "Top 10 algorithms in data mining", Knowl Inf Syst (2008) 14:137 DOI 10.1007/s10115-007-0114-2

[6] Huang Xiuchang, SU Wei, "An Improved K-means Clustering Algorithm", JOURNAL OF NETWORKS, 161-167, VOL. 9, NO. 1, JANUARY 2014.

[7] Ritu Yadav, Anuradha Sharma, "Advanced Methods to Improve Performance of K-Means Algorithm: A Review Clustering", Global Journal of Computer Science and Technology Volume 12, Issue 9, Version 1.0, 46-52, April 2012.

[8] Anupama Chadha, Suresh Kumar,"An Improved K-Means Clustering Algorithm: A Step For ward for Removal of Dependency on K", 2014 International Conference on Reliability, Optimization and Information Technology -ICROIT 2014, India, Feb 6-8 2014

[9] Anand Bahety, "Extension and Evaluation of ID3- Decision Tree Algorithm",11-18, ICCCS, 2014, ICCC, 2014.

[10] Vikas Chaurasia, et al, Carib.j., Early Prediction of Heart Diseases Using Data Mining Techniques, SciTech,2013,Vol.1,208-21