www.ijiarec.com

ISSN:2348-2079

Volume-7 Issue-1

International Journal of Intellectual Advancements and Research in Engineering Computations

Intelligent heart disease prediction system using data mining techniques

P.Dhivya[1], N.Naveenraj[2], K.Savitha[2], C. Vijayakumar[2], C. Vijayalakshmi[2]

[1]Assistant Professor Department of Information Technology

[2]UG Students, Department of Information Technology

Nandha Engineering College (Autonomous), Erode, Tamilnadu, India

ABSTRACT

Cardiovascular disease remains the cause of deaths worldwide. The percentage of premature death from this disease ranges from in high income countries and 42 % in low income countries. This shows the importance of heart predicting disease at the early stage. In 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 records are preprocessed initially using data mining techniques and then the attributes are classified using the Fuzzy C means classifier. In the classification stage 13 attributes are given as input to the Fuzzy C Means (FCM) classifier to determine a risk of heart attack. It is an unsupervised clustering algorithm, which allows one piece of the data to belong to two or more clusters. The proposed system will provide an aid for the diagnosis the disease in a more efficient way. The efficiency of the classifier is tested using the records collected from 1002 patients, which gives a classification accuracy of 94%. The result shows that the proposed clustering algorithm can predict the likelihood of patients and getting a heart attack in a more efficient and cost and effective way than the other well-known algorithms.

INTRODUCTION

IHDPS is a Web-based, user-friendly, scalable, reliable and expandable. A wide variety of areas including marketing, customer relationship, engineering, management, medicine, crime analysis, expert prediction, Web mining besides others utilizes Data mining. Numerous fields associated with the medical services like prediction of effectiveness of the surgical procedures, medical tests, medications, and discovery of relationships among clinical and the diagnosis data as well employ Data Mining methodologies. Providing precious services at affordable costs is a major constraint encountered by the healthcare organizations (hospitals, medical centers). IHDPS is a Web-based, scalable, user-friendly, and expandable. A wide variety of the areas including marketing, customer relationship management, engineering, medicine, crime analysis, expert prediction, Web mining, and mobile computing, besides others utilize Data mining. Numerous fields associated with medical services like prediction of effectiveness of surgical procedures, medical tests, medication, of relationships among clinical diagnosis data as well employ Data Mining methodologies. it providing precious services at affordable costs is a major constraint encountered by the healthcare organizations .

Working on heart disease patients databases is one kind of a real-life application. The detection of a disease from several factors or symptoms is the multi-layered problem and might lead to a false assumptions frequently associated with the erratic effects. It is appears reasonable to try utilizing the knowledge and experience of several specialists collected in a databases towards assisting the diagnosis process It is the possible for a healthcare

Author for correspondence:

Department of Information Technology Nandha Engineering College (Autonomous),Erode, Tamilnadu, India

industries to gain advantage of Data mining by employing the same as an intelligent and diagnostic tool. It is possible to predict the efficiency of medical treatments by building the Data Mining applications. Data mining can deliver an assessment of which courses of actions prove the effective by comparing and evaluating cause, symptoms, and courses of treatments.

METHODOLOGY

IHDPS uses to build the mining models. It consists of six major phases of business understanding, data understanding, data preparation, modeling, and deployment. Business understanding phase is focuses on understanding the objectives and requirements from a business perspective and converting. This knowledge into data mining problem definition, and designing a preliminary plan to achieve the objectives of normalization. Data understanding phase uses the raw the data and proceeds is understand the data, identify its quality, and detect interesting subsets to form hypotheses for a hidden information. Data preparation phase constructs final datasets that will be fed into the modeling tools. This includes table, record, and attributes selection as well as data cleaning and transformation. The modeling phase selects and applies various types of techniques, and calibrates their parameters to optimal values. The evaluation phase evaluates the model to ensure. It achieves the business objectives. The deployment phase specifies the tasks that are needed to use the models. Data Mining Extension (DMX), a SQL-style query language for data mining, is used for accessing the models' contents. Tabular and graphical visualizations are incorporated to enhanced analysis and interpretation of results.[1-4]

EXISTING SYSTEM

In Existing system are Very few systems uses the available clinical data for prediction purposes and even if they do, they are restricted by the large number of association rules that apply. Diagnosis of a condition solely depends upon the Doctors' intuition and patient's records. The National patient safety foundation cites that medical patients feel they have had experienced a medical error or missed diagnosis.

Drawbacks

- Detection is not possible at an earlier stage.
- Clinical decisions are often made based on doctors intuition and experience rather than on the knowledge rich data hidden in the database
- This practice leads to unwanted biases, errors and expensive medical costs which affect the quality of service province to patients
- There are many ways that a medical misdiagnosis can present itself. Whether a doctor is at fault, or hospital staff, a misdiagnosis of a serious illness can have very extreme and harmful effects

PROPOSED SYSTEM

Medical Misdiagnoses are a serious risk to a our healthcare profession. If they continue, then people will fear going to the hospital for treatment. The proposed system acts as decision support system and will prove to be an aid for a physician with the diagnosis. The algorithm, Fuzzy c means uses clustering and makes use of a clusters and data points to predict the relativity of an attribute. Each data point is associated with multiple. Clusters depending upon the membership degrees.

Advantages

- We proposed an integration of clinical decision support with a computer based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome.
- This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to the generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions
- Fuzzy c Means (FCM) is very flexible and is widely in various domains with high rates of success.

FUZZY C MEANS (FCM) CLASSIFIER

In fuzzy clustering, every point has a degree of belonging to clusters, fuzzy logic, rather than be longing 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 in any point x has a set of coefficients giving the degree of being in the kit cluster wk.(x). With fuzzy c-means (FCM), the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the clusters [5-8]

The degree of belonging, wk. (x), is related inversely to a distance from the cluster center as calculated on the previous pass. It is also depends upon a parameter m that controls how much weight is given to the closest center. Assign randomly to the each point coefficients for being in the clusters.

- Repeat until the algorithm has converged change between two iterations is no more than , the given sensitivity threshold) :
- Compute the centroid for using the formula above each cluster.
- For each point, compute its a coefficients of being in the clusters, using the formula above.
- The algorithm minimizes intra-cluster variance as well as , but has the same problems as k- means; the minimum is a local minimum, and results depend on the initial choice of weights.
- Using a mixture of Gaussians along with the EM Algorithm is a more statistically formalized
- Method which includes some of these ideas partial membership in classes.

MODULE

Registration module

In this module, registration for the multiple users. There are multiple owners, multiple attribute authorities, and multiple users. The attribute hierarchy of files – leaf nodes is atomic file categories while internal nodes are compound categories. Dark boxes are the categories that a personal domain's data reader has access to. There are Two Registration is here.

1. Doctor Registration
2. Patient Registration

For each registration their unique key and password automatically generated here. These two attributes used for their login.

Access control module

Access control has the login page for patients and doctor. They login individually to access their page.

Doctor modules contains

Appointment module

Appointment details of patients and their history and Problem. Doctor can click the button to be added to visited list and write prescription for the patients.

Dataset module

This module will contains the 1000 sample datasets .which used to clustering and find membership matrix of each particular dataset using fuzzy C means clustering algorithm contains 13 attributes of patients such as

Serial no	**Attribute Name**		**Attribute value**	**Attribute Condition**
1	Sex	Gender	Male/Female	Male/Female
2	Smoking	Smoking habit	Yes	Abnormal
			No	Normal
3	Sugar	diabetes	>120	Abnormal
			<120	Normal
4	Cholesterol	Serum cholesterol	0-160	Abnormal
			160-250	Normal
			250-410	Abnormal
5	Rest BP	Resting blood pressure	0-90	Abnormal
			90-192	Normal
			192-300	Abnormal
6	ECG	Resting electrographic results	Normal	Normal
			Having ST-T way abnormality	Normal
			Showing probable or definitleft ventricular hypertrophy	Abnormal
			NULL	NULL
7	THAL	thallic	Normal	Normal
			Fixed	Abnormal
			Reversible Defect	Abnormal
			NULL	NULL
8	Chest Pain	Chest pain type	Typical Type 1 Angina	Abnormal
			Typical Type Angina	Abnormal
			Non Angina	Normal
			Asymptomatic	Abnormal
			NULL	NULL
9	Angina	Exercise induced angina	Yes	Abnormal
			No	Normal
			NULL	NULL
10	St Slope	Slope of the peak exercise ST segment	Up Sloping	Abnormal
			Flat	Normal
			Down Sloping	Abnormal
11	Old Peak	ST depression induced by exercise relative to rest	Continuous	Abnormal
			Non-Continuous	Normal
			NULL	NULL
12	Vessels	Number of major vessels colored by fluoroscopy	0	Normal
			1	Abnormal
			2	Abnormal
			NULL	NULL
13	Heart Rate	Maximum heart rate achieved	<72	Abnormal
			72	Normal
			>72	Abnormal

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.. In the classification stage, a FCM classifier is used to classify the data as heart disease present or not. The results of classification experiment, preformed over data sets obtained from 1000 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.

FUTURE ENHANCEMENTS

In future, further enhancements of these techniques, e.g., Time Series, clustering and Association Rules. Continuous data can also be used instead of just categorical data. Another area is to use Text mining to mine the vast amount of unstructured data available in healthcare databases. Another challenge would be to integrate data mining and text mining.

REFERENCES

[1]. PHP & MySQL Novice to Ninja – by Kevin Yank
[2]. Head First PHP & MySQL – by Lynn Beighley & Michael Morrison
[3]. PHP & MySQL Web Development – by Luke Welling & Laura Thompson
[4]. PHP & MySQL: The Missing Manual – by Brett McLaughlin
[5]. PHP: A Beginner's Guide – by Vikram Vaswani
[6]. Data Communications and Networking, by Behrouz A Forouzan.
[7]. Computer Networking: A Top-Down Approach, by James F. Kurose
[8]. Blake, C.L., Mertz, C.J.: "UCI Machine Learning Databases", http://mlearn.ics.uci.edu/databases/heartdisease/, 2004.