ISSN 2395-1621

Intelligent Disease Predicting System

#1Aishwarya P. Aundhkar, #2Priyanka S. Chavan, #3Swapnali U. Dubal, #4Shital A. Padwal, #5Smita S. Pawar

12345Research Student, Department of Computer Science and Engineering, Arvind Gavali College of Engineering, Panmalewadi, Varye, Satara-415015, Maharashtra, India.

#6Prof. Karande H.A.
#6Assistant Professor, Department of Computer Science and Engineering, Arvind Gavali College of Engineering, Panmalewadi,Varye, Satara-415015, Maharashtra, India.

ABSTRACT

This paper presents the concept of clustering algorithm based data mining for medical environment. Data mining is the process of extracting useful data from huge amount of data. The main purpose of data mining is to provide superior results. Data mining mainly used in medical field to predict disease such as heart disease, cancer etc. By using clustering algorithm, we can easily analyze huge amount of data. In clustering we implementing K-means clustering algorithm.

Keywords: Data Mining, K-means Clustering algorithm.

ARTICLE INFO

Article History

Received: 25th March 2017

Received in revised form : 25th March 2017

Accepted: 25th March 2017

Published online : 4th May 2017

I. INTRODUCTION

In medical environment, huge amount of information is available but to analysis this information knowledge is weak. To identify hidden relationship in data there is lack of data analysis tool. Data mining is the concept of finding different patterns or co-relations in the large relational database. In data mining disease predicting plays an important role. Data mining deals with knowledge discovery in database. This paper analyzes the heart disease predictions using K-means clustering algorithm. Medical data mining technique like clustering algorithm is implemented to analyze the different kinds of heart based problems. Clustering algorithm like K-means are the data mining techniques used in the medical field. By using this technique, the level of heart disease can be predicted.

II. LITERATURE REVIEW

Shadab Adam Pattekari et al. (2012)[2] explain Data mining mainly used in medical field to predict disease such as heart disease, cancer etc. V. Manikantan et al.(2013) [1] explain Mining and structure methods which are present in medical data mining are used to analyze medical data and its content. By using K-means clustering algorithm the heart disease database is clustered, which will fetch the data applicable to heart attack from the database. Jyoti Soni et al.(2011)[3] explain In medical environment, huge amount of information is available but to analysis this information knowledge is weak. To identify hidden relationship in data there is lack of data analysis tool. Sellappan Palaniappan et al.(2008)[4] explain the health care data which is collected from different health care industry which is unfortunately not useful to identify hidden relationship to take effective decision.using medical data such as age, sex, weight, blood pressure, sugar and chest pain etc. it can predict the level of heart attack of patient. Shadab Adam Pattekari et al.(2012)[5] explain traditional support system cannot answer complex queries for diagnosing heart disease so these system can give the intelligent clinical decisions for heart disease.

III. METHOD

Clustering is the task of grouping a set of objects in a such way that objects in the same group are more similar to each other.
K-means clustering aims to partition n observation into K cluster in which each observation belongs to the cluster with the nearest mean. The result in a partition of the dataset.

K-means Algorithm

Step1: Select initial center cluster as a K-points called “means”.
Step2: Each dataset point from database is assigned to its closed cluster based on the distance between dataset point and cluster.

Step3: Recomputed the initial center cluster to calculate average of cluster.
Step4: Repeat Step 2 and 3 until the cluster coming closer.

Example

Suppose following are the given dataset-

(5,10) , (6, 8), (4,5) , (7, 10), (8,12) , (10,9), (12,11), (4,6)

From above dataset point initial center clusters are:-
(5, 10), (7,10), (12,11)

Dataset points are assigned to closed cluster depend upon distance between dataset point and cluster.

The distance is calculated using following formula:-
P(a,b) = (x2 - x1) + (y2 - y1)

Where
(x1,y1) are dataset point
And
(x2,y2) are initial cluster center
Here,
(1)
(x1,y1) = (5,10)
(x2,y2) = (5,10)
By using distance formula,
P(a,b) = (x2 - x1) + (y2 - y1)
= (5-5) + (10-10)
= 0

(x1,y1) = (5,10)
(x2,y2) = (7,10)
By using distance formula,
P(a,b) = (x2 - x1) + (y2 - y1)
= (7-5) + (10-10)
= 2

(x1,y1) = (5,10)
(x2,y2) = (12,11)
By using distance formula,
P(a,b) = (x2 - x1) + (y2 - y1)
= (12-5) + (11-10)
= 8

The dataset point (5,10) have a minimum distance 0 so, it belongs to initial cluster (5,10).
In this way all dataset point are assigned to its closed cluster. Cluster and there dataset point are shown in following table.

Cluster 1 (5,10)	Cluster 2 (7,10)	Cluster 3 (12,11)
(5,10)	(7,10)	(10,9)
(6,8)	(8,12)	(12,11)
(4,5)	-	-
(4,6)	-	-

Table 1: clustered dataset

Recomputed new cluster by taking mean of all dataset points in cluster.
Recomputed using following formula-
(x1+x2+x3+….+xn)/n
(y1+y2+y3+….+yn)/n=(x,y)
Where n= number of dataset point
C1 = (5+6+4+4)/4, (10+8+5+6)/4
=(4.75,7.25)
C2=(7+8)/2, (10+12)/2
=(7.5,11)
C3=(10+12)/2, (9+11)/2
=(11,10)

Center of new clusters are
C1 = (4.75,7.25)
C2 = (7.5,11)
C3 = (11,10)

By using above centers following are the new clusters

Cluster 1 (4.75,7.25)	Cluster 2 (7.5,11)	Cluster 3 (11,10)
(5,10)	(7,10)	(10,9)
(6,8)	(8,12)	(12,11)
(4,5)	-	-
(4,6)	-	-

Table 2: Recomputed clustered dataset

We are using following parameters for clustering dataset

1. age:age in years
2. sex:sex(1 = male;0 = female)
3. cp:chest pain type
   - Value 1:typical angina
   - Value 2:atypical angina
   - Value 3:non-anginal pain
   - Value 4:asymptomatic
4. trestbps: resting blood pressure(in mm Hg on admission to the hospital)
5. chol:serum cholestoral in mg/dl
6. fbs(fasting blood sugar>120mg/dl) (1 = true; 0 = false)
7. restecg:resting electrocardiograhic results
   - Value 0: normal
   - Value 1: having ST-T wave abnormality(T wave inversions and/or ST elevation or depression of> 0.05 mV)
   - Value 2:showing probable or definite left ventricular hypertrophy by Estes' criteria
8. thalach: maximum heart rate achieved
9. exang: exercise induced angina(1 = yes; 0 = no)
10. oldpeak = ST depression induced by exercise relative to rest
11. slope: the slope of the peak exercise ST segment
    - Value 1: upsloping
    - Value 2: flat
    - Value 3: downsloping
12. ca: number of major vessels(0-3) colored by fluoroscopy
13. thal:3 = normal;6 = fixed defect;7 = reversible defect
14. num: diagnosis of heart disease(angiographic disease status)
    - Value 0: <50% diameter narrowing
    - Value 1: >50% diameter narrowing

(in any major vessel: attribute 59 through 68 are vessels)

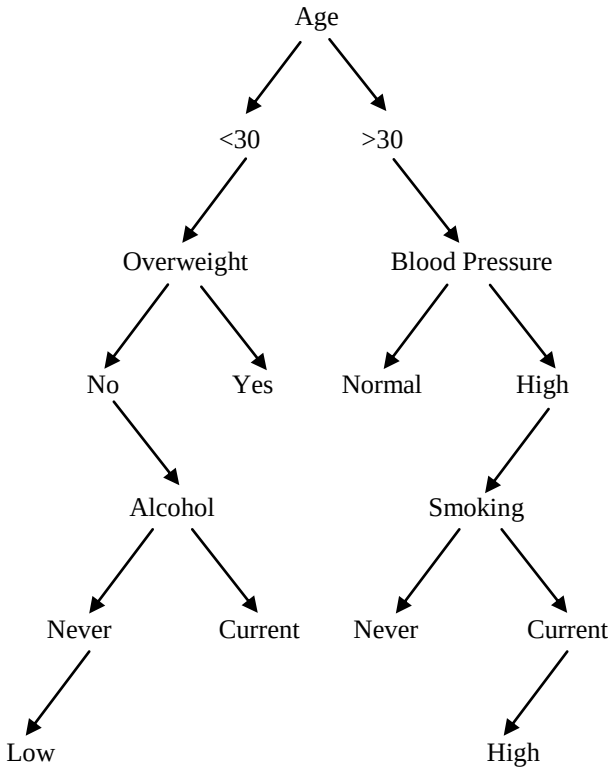
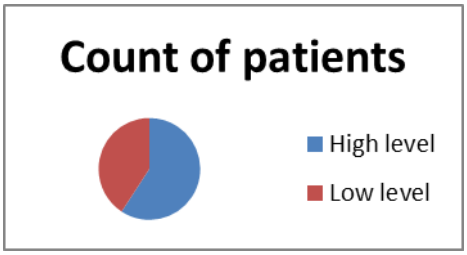


Fig 1: Decision tree for prediction of heart attack level

IV. RESULTS

Suppose, we have 303 record of patient to analyze how many patient are in low level and high level of heart disease.Then this detail of patient are given to the K-means clustering algorithm. After this, clustering algorithm distrubute the record in low level and high level of heart disease.After that it will give the result that is ,

Low level patient are=180
High level patient are=123



Discussion

Based on the result, it was found that it required the dataset related to medical field to analyze the level of heart disease. In that dataset it must have parameters which are given to K-means clustering algorithm.

V. CONCLUSION

In this paper we have proposed K-means clustering algorithm for medical data analysis. In proposed paper it analyzes number of records related to heart disease and it will show result within a fraction of second that heart disease is high or low. In future, we can easily analyze different disease like cancer, hepatitis, diabetes.

References

[1] V. Manikantan and S. Latha, "Predicting the analysis of heart disease symptoms using medicinal data mining methods", International Journal of Advanced Computer Theory and Engineering, vol. 2 ,pp.46-51,2013.

[2] Shadab Adam Pattekari and Alma Parveen," Prediction system for heart disease using naïve bayes", International Journal of Advanced Computer and Mathematical Sciences, vol.3,pp 290-294,2012.

[3] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni "Predictive data mining for medical diagnosis: an overview of heart disease prediction" International Journal of Computer Science and Engineering, vol. 3 ,2011

[4] Sellappan Palaniappan Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, vol.8 No.8, August 2008

[5] Shadab Adam Pattekari and Asma Parveen, "Prediction System For Heart Disease Using Naive Bayes", International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624. vol 3, Issue 3, 2012, pp 290-294

