

Comparative Analysis of K-Means Algorithm in Disease Prediction

K.Rajalakshmi¹, Dr.S.S.Dhenakaran², N.Roobini³

Abstract— an extremely fastest growing field is medical science. The modern medicine system generates huge amount of data every day. Without the computing of these data is totally waste. To turns these data into useful pattern, the data are needed to be mined. The medical data mining are useful to produce efficient results on prediction based system in medical oriented fields. This paper analysis various diseases prediction using K-means algorithm.

Index Terms— data mining, K-means algorithm, medical data mining

I. INTRODUCTION

Data mining is the knowledge discovering technique that used to find the hidden values from enormous amount of data. As the patients population and treatments increases the medical databases also growing every day. The transactions and analysis of these medical data is complex without the computer based analysis system. The computer based analysis system provides the automated medical diagnosis system. This automated diagnosis system support the medical practitioner to make efficient decision in treatment and disease prediction. Data mining is the rapidly growing areas for the doctors to handling the huge amount of patient's data sets in many ways such as make sense of complex diagnostic tests, interpreting previous results, and combining the different data together. Traditionally Infirmaries decision is shaped by the medical practitioner's observations and forecasting the knowledge rather than the knowledge which obtain from the large amount of data. This automated diagnosis system leads to increases the quality of service provided to the patients and decreases the medical expenditure.

II. DATAMINING

Data mining is the process of combining the different data source and derives the new pattern from that data collection. The following diagram represents different stages of data mining process. The data mining process uses different techniques .In these paper we have presented clustering techniques in data mining

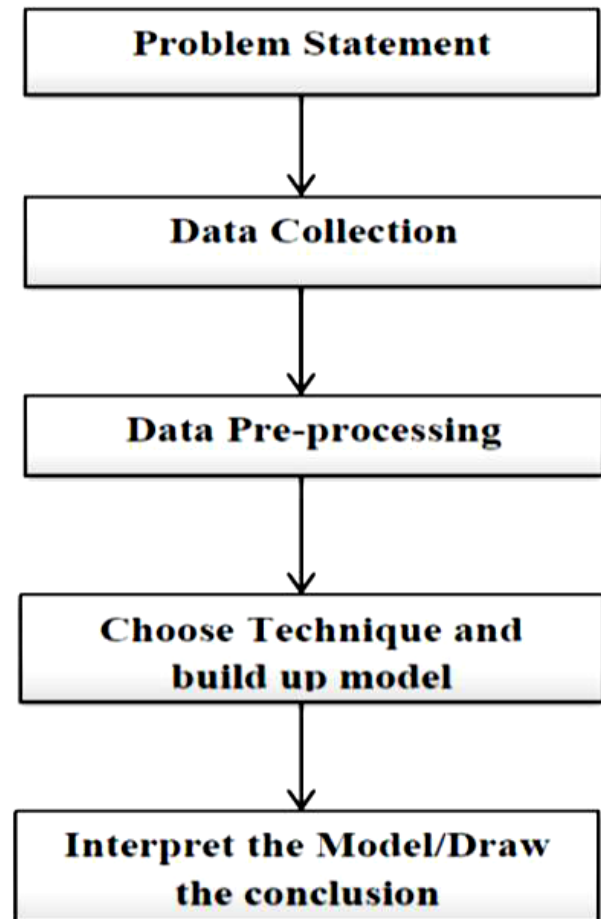


Fig.1 Data mining process

III. CLUSTERING

Clustering is the one of the machine learning technique. This is unsupervised classification which has no proper predefined classes. Clustering is the process of segmenting set of data into meaningful sub-class called clustering. The stages of clustering method can be explained as follows:

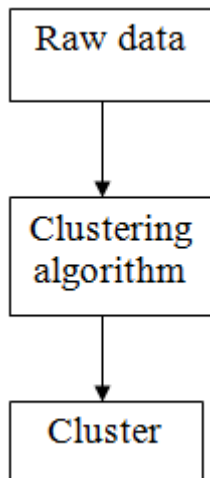


Fig.2 Stages of clustering

These clustering methods can be classified as follows:

- a) Partitioning method
- b) Hierarchical method
- c) Density-based method
- d) Grid-based method
- e) Model-based method
- f) Constraint-based method

a) Partitioning method

In partitioning clustering data objects are segmented into several subsets. Each and every object presented in exactly in one subset.

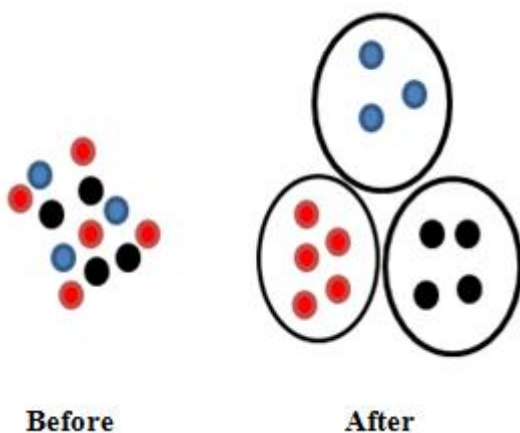


Fig.3 Partitioning clustering

b) Hierarchical clustering method

In hierarchical clustering the datasets are segmented as hierarchy of clusters. This method uses connectivity approach. These hierarchy cluster created by using agglomerative and divisive algorithms.

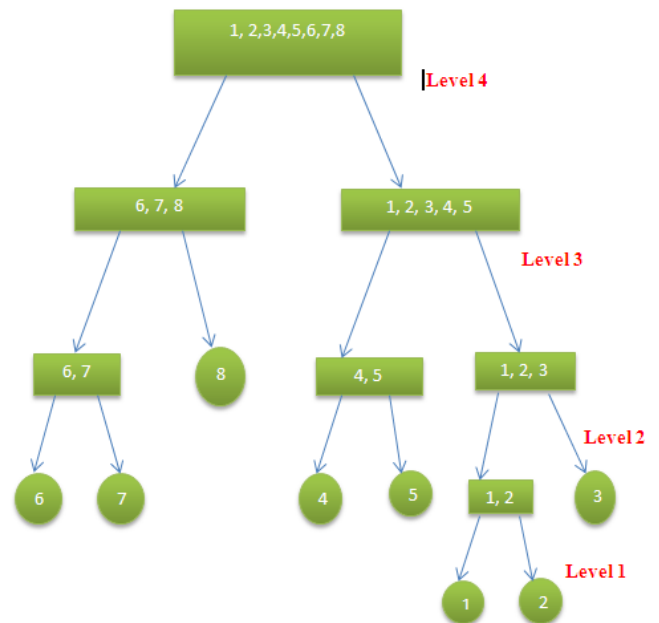


Fig.4 Hierarchical clustering

c) Density based clustering method

This density based clustering method is find the clusters based on the density of data points in region.

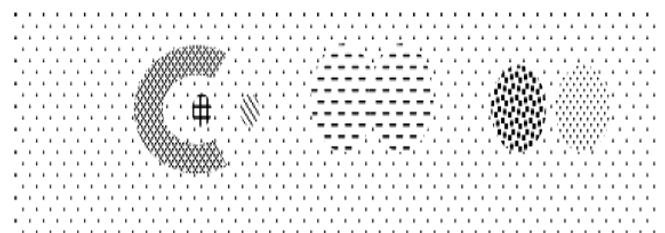


Fig.5 Density based cluster

d) Grid based clustering method

In grid based clustering the data space are divided into finite number of cells it forms the grid structure and performing clustering on grid.

IV. K-MEANS ALGORITHM

K-means clustering is partition based clustering technique. It was introduced by Macqueen in 1967. This is one of the simplest forms of unsupervised learning algorithm that used to solve cluster based problem. In K-means clustering

algorithm given data sets are divided into fixed number of clusters. This algorithm consists of two phases there are as follows:

1. Fix the K value in advance and choose K center in random manner.
2. Assign data objects to nearest center

i) Method:

The process of K -Means algorithm as follows:

INPUT: Number of desired clusters K and a database $D = \{d_1, d_2 \dots d_n\}$ Containing n data objects.

OUTPUT: A set of K clusters.

ii) Steps:

1. Determine the center of clusters
2. Determine the distance of each data object to the centroids
3. Set the position of each cluster to mean of all data points belonging to that cluster
4. Repeat steps 2&3 until convergence

K -means algorithm uses iterative method that minimizes the sum of distances from each object to its cluster centroid. These algorithm processes until the sum of distance cannot be decreased further.

ii) Advantages of k -means algorithm:

- a) The main advantage of this algorithm is simplicity and its speed which allows running large datasets.
- b) K -Means may be faster than hierarchical clustering (if K is small).
- c) K -Means may produce tighter clusters than hierarchical clustering, especially if the clusters are globular.

V. COMPARISON OF DIFFERENT PREDICTION USING K -MEANS CLUSTERING

Data mining techniques are mainly used in predication based computing. In data mining there wide range of algorithms are available. Normally K -means algorithm is used in disease predication system. Here same of disease which are predicated using K -means algorithm are discussed below with results. Diseases which are predicated are as follows:

- Heart Disease
- Diabetics
- Liver disease
- Cancer

There are many research topics available for heart disease prediction using K -means algorithm. The following table

presents comparison of disease, data mining techniques and the accuracy of the data mining techniques

| ALGORITHM | ACCURACY | DISEASE |
|--------------------------------------|----------|---------------|
| K-Means | 98.24 | Heart disease |
| K-Means | 78 | Diabetics |
| K-Means(Attribute Based) | 80.198 | Heart Disease |
| K-Mean based MAFIA | 74 | Heart Disease |
| K-Mean base MAFIA with ID3 | 85 | Heart Disease |
| K-Mean based MAFIA with ID3 and C4.5 | 92 | Heart Disease |
| K-Means | 78 | Diabetics |

Table 1. K-means with different disease

K -means algorithm may be used alone and with different techniques in data mining like Naïve Bayes, decision tree and with other classification technique.

VI. CONCLUSION

We presented this paper to analyze the various disease prediction applications in the healthcare domain to discover new range of pattern information using k -means clustering. This data mining based prediction system are reduces the human effects and cost effective one.

REFERENCE

- [1] V.Bala Sundar, T.devi, N. saravanan "Development of a Data Clustering Algorithm for Predicting Heart", International Journal of Computer Applications (0975 – 888) Volume 48– No.7, June 2012.
- [2] Marina Gorunescu, Florin Gorunescu, Radu Badea, and Monica Lupșor "Evaluation on liver fibrosis stages using the k -means clustering algorithm", Annals of University of Craiova, Math. Comp. Sci. Ser. Volume 36(2), 2009, Pages 19{24 ISSN: 1223-6934.
- [3] VelidePhani Kumar and Lakshmi Velide," Data Mining Approach for Prediction and Treatment Of diabetes Disease" VelidePhani Kumar-et al., IJSIT, 2014, 3(1), 073-079
- [4] Atul Kumar Pandey, Prabhat Pandey, K.L. Jaiswal, Ashish Kumar Sen ," DataMining Clustering Techniques in the Prediction of Heart Disease using Attribute Selection Method", International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 10, October 2013
- [5]. G.Visalatchi, S.J Gnanasoundhari, Dr.M.Balamurugan," A Survey on Data Mining Methods and Techniques for Diabetes Mellitus", G.Visalatchi et al, International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 2, February- 2014, pg. 100-105
- [6] Sathyabama Balasubramanian, Balaji Subramani," Symptom's Based Diseases Prediction In Medical System By Using K-Means Algorithm", International Journal of Advances in Computer Science and Technology Available Online at <http://warse.org/pdfs/2014/ijacst13322014.pdf>
- [7] V. Manikantan & S. Latha , "Predicting the Analysis of Heart Disease Symptoms Using Medicinal Data Mining Methods", International Journal on Advanced Computer Theory and Engineering (IJACTE) ISSN (Print) : 2319 – 2526, Volume-2, Issue-2, 2013.
- [8] K. Rajalakshmi Dr. S. S. Dhenakaran," Analysis of Data mining Prediction Techniques in Healthcare Management System", Volume 5, Issue 4, April 2015 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering