SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING

VOLUME 4, ISSUE 2
ISSN 2054-7412

A Comparison of Different Clustering Methods for MIT BIH ECG Data

[1]Sharathchandra Chilkuri , [2]M. Prabhakara Reddy, and [3]Ibrahim Patel ,
[1,2]*Dept. of Biomedical Engineering, B.V. Raju Inst. of Tech., Narsapur Medak,(T. S) India;*
[3]*Dept. of ECE B.V. Raju Inst. of Tech., Narsapur Medak,(T. S) India*
sharathchandrachilkuri@gmail.com; prabhakarareddy.m@bvrit.ac.in; ptlibrahim@gmail.com;

ABSTRACT

Electrocardiogram can be occasionally or continuously measured from living Human beings. Regardless of their Disease, Now a day's physicians or doctors are suggesting to take ECG. These signals reflect the physiological processes and electrical activity of the Heart. Therefore, the study of ECG signals is essential for both medical applications and scientific studies for this purpose one requires best clustering method. It is difficult to provide a best clustering methods for the ECG signals because these categories may overlap, so that a method may have features from several categories. Nevertheless, it is useful to present a relatively organized picture of the different clustering methods.

Keywords: ECG; Clustering; MITBIH; QRS Detection; Filtering.

1 Introduction

The Electrocardiogram signal is generated by polarization and depolarization of the heart that occurs when pumping blood throughout the human body, and it can be recorded by contacting electrodes to the skin at specific locations on the body. It provides the valuable information regarding the cardiovascular diseases. Any abnormality in rhythm can provide useful information about the type of disease. in ecg QRS complex is a dominant of electrocardiographic signal. Its amplitude and time analysis, shape and appearance time of adjacent rhythms estimation can be used to diagnose a wide range of heart diseases. QRS complex is necessary for the determination of the heart rate, and as reference for beat alignment.Thus, the obvious problem is the precise definition of the occurrence time and other various parameters of QRS-complex.

Various methods for classification of arrhythmias have been developed by researchers and clustering technique is one of them. Although it is an unsupervised type of technique, it is advisable technique for analysis and interpretation of long term ECG Holter records. In this paper, we are tesing four clustering has been used for analysis.

2 Clustering

Clustering [1] can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way".A *cluster* is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters.

DOI: 10.14738/aivp.42.2008
Publication Date: 25rd April, 2016
URL: http://dx.doi.org/10.14738/aivp.42.2008

2.1 K-Means Algorithm

K-means (MacQueen, 1967)[1] is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (1)$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centers.

2.2 K-Medoids Algorithm

The K-means algorithm [1] is sensitive to outliers since an object with an extremely large value may substantially distort the distribution of data. How might the algorithm be modified to diminish such sensitivity? Instead of taking the mean value of the objects in a cluster as a reference point, a Medoid can be used, which is the most centrally located object in a cluster. Thus the partitioning method can still be performed based on the principle of minimizing the sum of the dissimilarities between each object and its corresponding reference point. This forms the basis of the K-Medoids method. The basic strategy of K Medoids clustering algorithms is to find k clusters in n objects by first arbitrarily finding a representative object (the Medoids) for each cluster. Each remaining object is clustered with the Medoid to which it is the most similar. K-Medoids method uses representative objects as reference points instead of taking the mean value of the objects in each cluster. The algorithm takes the input parameter k, the number of clusters to be partitioned among a set of n objects

2.3 Hierarchal Algorithm

These methods construct the clusters by recursively partitioning the instances in either a top-down or bottom-up fashion. These methods can be subdivided as following:

- Agglomerative hierarchical clustering — each object initially represents a cluster of its own. Then clusters are successively merged until the desired cluster structure is obtained.
- Divisive hierarchical clustering — All objects initially belong to one cluster. Then the cluster is divided into sub-clusters, which are successively divided into their own sub-clusters. This process continues until the desired cluster structure is obtained.

The result of the hierarchical methods is a dendrogram, representing the nested grouping of objects and similarity levels at which groupings change. A clustering of the data objects is obtained by cutting the dendrogram at the desired similarity level. The merging or division of clusters is performed according to some similarity measure, chosen so as to optimize some criterion (such as a sum of squares). The hierarchical clustering methods could be further divided according to the manner that the similarity measure is calculated

2.4 Fuzzy C-Means Algorithm

The most popular fuzzy clustering algorithm is the fuzzy c-means (FCM) algorithm. Even though it is better than the hard K-means algorithm at avoiding local minima, FCM can still converge to local minima of the squared error criterion. The design of membership functions is the most important problem in fuzzy clustering; different choices include those based on similarity decomposition and centroids of clusters. A generalization of the FCM algorithm has been proposed through a family of objective functions. A fuzzy c-shell algorithm and an adaptive variant for detecting circular and elliptical boundaries have been presented. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunn in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad , 1 \le m < \infty \qquad (2)$$

where *m* is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster *j*, x_i is the *i*th of d- dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}} \quad , \qquad c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

This iteration will stop when $\max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \varepsilon$, where ε is a termination criterion between 0 and 1, whereas *k* are the iteration steps. This procedure converges to a local minimum or a saddle point ofJ_m.

3 Implementations

Step 1: Load the MIT-BIH ECG Database

Step 2: Convert the MIT-BIH ECG Database into MATLAB Readable Format

Step 3: Applied the IIR Butterworth filter along with notch filter on ECG Database

Step 4: The absolute slope [4] i.e. absolute value of the difference between two consecutive samples is calculated to enhance the signal in the region of QRS-complex. The absolute value of slope of the ECG signal is used as an important discriminating feature because absolute slope of the signal is much more in the QRS-region than in the rest of the region. Fig. shows the absolute slope of the filtered ECG signal.

Step 5**:** The various steps of four clustering algorithm four clustering algorithms as described in above section are followed in order to find the two cluster centers namely the QRS-cluster centre and the non QRS-cluster centre.

Step 6: After finding two cluster centers using four clustering algorithms, the slope curve shown in Fig. is scanned. The membership of slope, at a given sampling instant, is found. An output is 2 if a sample belongs to a QRS-cluster and output is 1 if it belongs to a non-QRS-cluster

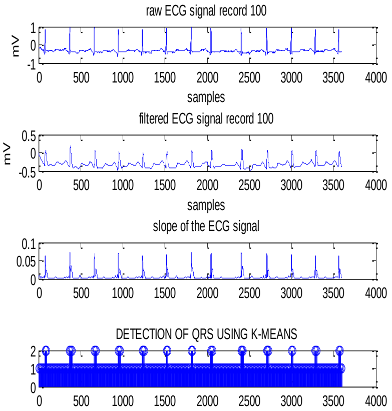

Figure 1: Implementation of K-MEANS

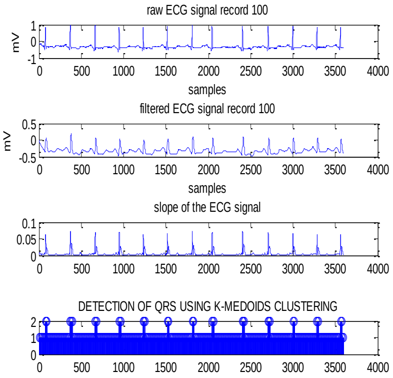

Figure 2: Implementation of K-MEDOIDS

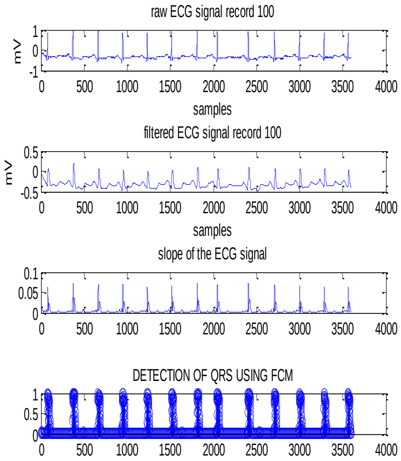

Figure 3: Implementation of FUZZY C-MEANS

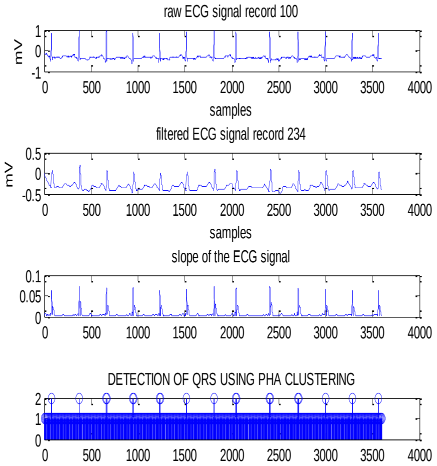

Figure 4:Implementation of HIERARICHAL

4 Result Analysis

In this paper we implemented four algorithms on MITBIH ECG Data. The four algorithms compared by four parameters i.e. Sensitivity, Specificity [6], Predictivity, Speed.

The above mentioned algorithms not only detect the QRS complexes of ECG, but also delineate them accurately detection is said to be true positive (TP) if the algorithm correctly discerns the QRS-complex and it is said to be false negative (FN) if the algorithm fails to detect the QRS complex. False positive (FP) detections are obtained if non QRS-wave is detected as a QRS-complex. The ECG signals used for analysis and detection in this work are a part of MIT-BIH Arrhythmia Database given on the website of MIT-BIH. The said algorithm is applied on total of 48 records from database.

It is observed that, in the case of normal beats (i. e. for record number 100, 101, 102, 104, 105, 106, 107, 112, 113, 115, 117, 119, 121, 122, 123, 201, 202, 209,212, 213, 215, 217, 219, 220, 221, 222, 223, 228, 230, 231, 232, 234) and right bundle branch block (i.e. for record numbers118,124),[7] the results are encouraging and almost all the beats were detected successfully. Similarly, in the case of left bundle branch block also (i.e. for record numbers 111,207, 214), the total number of complexes detected are accurate and percentage range of Se and P+ is satisfactory. As the algorithm has been implemented in MATLAB working environment, therefore the part of the whole signal of each data set has been operated. In order to evaluate the accuracy of detection of QRS complex, three essential parameters: sensitivity Se and the positive predictivity P+(detection rate), specificity are used as listed in Table 1 These parameters describe the overall performance of the detector and their values are calculated as follows

$$\text{Sensitivity} = TP / (TP + FN)$$

$$\text{Predectivity} = TP / (TP + FP)$$

$$\text{Specificity}=TN / (TN+FP)$$

Using the above formula, Table 1 clearly shows the results of four methods in terms of sensitivity, specificity, predictivity is obtained for all 48 MIT-BIH Records. Also the percentage of false positive detection and false negative detection for all records are very less.

Table 1: Comparison of Different Algorithms

Clustering Method	Sensitivity	Specificity	Predictivity	Time(Sec)
K-Means	100%	100%	97%	0.41-0.44
K-Medoids	100%	100%	96%	0.81-1
Hierarchical	100%	100%	98.39%	0.22-0.23
Fuzzy C Means	100%	100%	96.39%	0.19-0.22

5 Conclusion

The four clustering methods have been comprehensively tested using the MIT BIH database covering wide variety of QRS complexes.

In above discussed four methods Sensitivity and specificity is approximately 100%

All Four methods got more than 95% detection rate or Predictivity.

It is observed that hierarchal algorithm is suitable than K-Means, K-Medoids and FCM algorithm based on Predictivity for the data sets in MITBIH data base.

6 Future Scope

The project outcome is QRS and Non QRS clustering and comparison of various clustering methods. In future one can extend this project to multiple clusters like P, QRS, T, U waves. The information obtained from the above methods can be useful for ECG interpretation and analysis. For example Estimation of heart rate, HRV.

It is also possible to extend these methods for automatic annotation of ECG signal and diseases diagnosis not only for the ECG data but we can also use for other biological data such as radiological images, EEG, EMG, Genes data.

REFERENCES

[1]. A Tutorial on Clustering Algorithms; home.deib.polimi.it/matteucc/clustering/tutorial_html/index.html

[2]. www.physionet.org/physiobank/database/mitdb MIT Database.

[3]. Ms. Ananya, Dr.S.L.Nalbalwar,Swarali Seth, “Detection of QRS Complexes in ECG using Kmeans Algorithm” International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 5, May – 2014

[4]. S. S. Meht. "Development of FCM based algorithm for the delineation of QRS-complexes in Electrocardiogram", 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), 12/2009

[5]. Sharathchandra C, M.Prabhakara reddy, Ibrahim patel, Rameshwar “Identification of QRS complexes using Hierarichal clustering Algorithm” IJIRCCE, Volume 4, Issue 2, Feb-2016

[6]. Sharathchandra C, M.Prabhakara reddy, Ibrahim patel, Rameshwar “Mark out of QRS complexes using fuzzy C-means Algorithm” IJAREST, Volume 3, Issue 2, Feb-2016

[7]. J. A. Hartigan and M. A. Wong (1979) "A K-Means Clustering Algorithm", Applied Statistics, Vol. 28, No. 1, p100-108

[8]. B.U. Kohler, C. Hennig, and R. Orglmeister, “The principles of software QRS detection,” *IEEE Eng Biol. Mag*, vol. 21, pp. 42–57, 2002.

[9]. Matlab help, MATLAB MATHWORKS. http://www.mathworks.com

[10]. G.V. Van, K.V. Podmasteryev; Review of Algorithms Detection the QRS Complex based on machine Learning.

[11]. Jalil, Bushra, Olivier Laligant, Eric Fauvet, and Ouadi Beya. "Detection of QRS complex in ECG signal based on classification approach", 2010 IEEE International Conference