

Cardiac Abnormalities Detection from Compressed ECG using Principal Components Analysis (PCA)

Padmini. Bhat

M.Tech. 4th Sem., (DE&C),

Srinivas School of Engineering, Surathkal, Manglore,

p_bhat20@yahoo.com

Abstract- In telecardiology applications, ECG signals are compressed before transmission for the faster data delivery and to reduce the bandwidth. However, ECG analysis and diagnosis algorithms are applied on the original signal. Therefore, compressed ECG data needs to be decompressed first. Decompression will cause delay on the doctor's mobile device. This is undesirable in body sensor networks (BSNs) as high processing involved in decompression will waste valuable energy in the resource and power constrained sensor nodes. In this paper, in order to diagnose cardiac abnormality such as Ventricular tachycardia, the ECG signal is compressed using Karnaugh-Loeve algorithm. The feature extraction is done using PCA, and k-mean for clustering of normal and abnormal ECG signals.

Index terms: KL algorithm, principal component analysis(PCA), k-mean algorithm

I. INTRODUCTION

Electrocardiogram (ECG) signal has been intensively used by cardiac specialists to effectively diagnose cardiovascular diseases[1]. A typical ECG signal as shown in Fig. 1 contains special waves such as P, T waves as well as QRS complex.

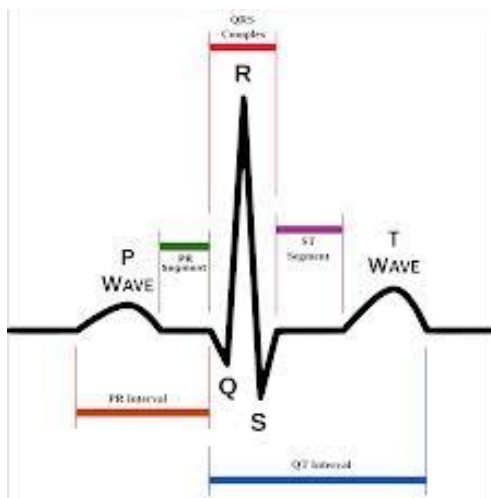


Fig-1 ECG waves

In wireless telemonitoring scenarios digitized ECG data need to be transferred as fast as possible and data must be compressed first to make the transmission energy efficient.

The abnormal cardiac condition considered in this paper is Ventricular Tachycardia. To diagnose the disorder, here first applied a lossless compression method such as karnaugh-loeve algorithm described in [2] before transmission. The compressed ECG signal is analysed and important features are extracted from it using Principal Component Analysis (PCA) [4]. The extracted features are classified as normal and abnormal using k-means algorithm [7], [8].

II. DATA COMPRESSION

The data is collected from publicly available ECG of MIT BIH Arrhythmia Database. The calculation of the KLT[2], [3] is typically performed by finding the eigenvectors of the covariance matrix. On its own, an orthonormal transformation does not effect data compression. The blocks of pixels are simply transformed from one set of values to another and, for reversible transformations, back again on reconstruction. To reduce the number of bits for representing an image, the coefficients are quantized, incurring some irreversible loss, and then encoded(run-length encoding) for more efficient representation. By decorrelating the data before these steps using the KLT, more data compaction can be achieved. With the pixel values forming the axes of a vector space, a rotation of this space can remove this correlation. The basis vectors of the new space define the linear transformation of the data. The basis vectors of the KLT are the eigenvectors of the image covariance matrix. Its effect is to diagonalize the covariance matrix, removing the correlation of neighboring pixels.

III. ANALYSIS OF COMPRESSED ECG SIGNAL

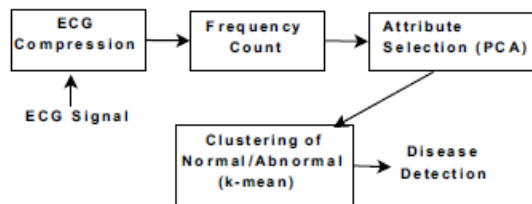
The compressed ECG data contains the characters set shown in Fig. 2. Character frequency calculation is performed for each compressed ECG segment. But attributes are large in number for clustering (normal and abnormal). Therefore, an attribute selection technique is applied called Principle Component Analysis (PCA)[4], [5] for dimensionality reduction.

```
@CSY&@0bQZwAA_~[~]
EAEwDE I@%&()'+,./:;<?
|g,Ab0aAAAAAEEIIMBN
00000000Y#paaapeel
T00000yabodefghijklmnop
qrstuvwxyzABCDEFGHIJK
LMNOPQRSTUVWXYZl
```

Fig 2 character set of compressed ECG data

IV. ATTRIBUTE SUBSET SELECTION

The PCA algorithm will generate a new small set of artificial variables called Principle Components which can be selected and fed to clustering system as shown.



A preprocessing of data using attribute selection algorithm is a critical issue in data mining solutions, since the training will be hard and inaccurate using large number of attributes. Also, it will make the system more complicated and the processing time will be large if the number of attributes keep increasing. In this paper, PCA algorithm is adopted which is appropriate if there is a set of samples with large number of variables (attributes). The algorithm will generate a new small set of artificial variables called Principle Components which can be selected and fed to clustering system.

By applying PCA on this data set, the covariance matrix of the data is generated. Next, eigenvectors and eigenvalues are derived for the covariance matrix which is then rearranged as a new matrix starting with the eigenvector that corresponds to the highest eigenvalue, and so on. As a result, this matrix will be $(n \times n)$ matrix where n is the number of variables (i.e in this case $n = 148$). After this, the scores matrix which is a $(n \times m)$ matrix is calculated where n is the number of samples and m is the number of variables. Equation 1 shows the general form to calculate the scores for the first principle component.

$$C1 = b11(X1) + b21(X2) + \dots + bp1(Xp) \tag{1}$$

where

- $C1$ = the sample score on the principal component 1
- $bp1$ = the regression coefficient (or weight) for observed variable p ,
- Xp = the sample value of variable no p

Similarly, other principal components (i.e. PC2, PC3, PC4, and so on) can also be calculated the first few components represent the high portion of data, which is clearly shown in Table I with the eigenvalues and the proportion of each eigenvalue of the total data. In table I it is clearly noticed that the first and second eigenvalues represent approximately 70% of the total data. Proportions of each eigenvalue in this table is derived by dividing the eigenvalue over the total summation of all eigenvalues obtained as in equation 2.

$$P_i = \frac{e_i}{\sum_{k=1}^m e_k}$$

P_i is proportion of the i th eigenvalue e_i is the i th eigenvalue M number of eigenvalues which is the same number of variables.

By taking just the first two principal components and their corresponding score will be used as an input for k-mean clustering part[6], [7], [8] to classify abnormal and normal ECG segments.

Principal Components	Eigenvalue	Proportion
Principal Component 1 (PC1)	82.56614	0.56552
Principal Component 2 (PC2)	20.04823	0.13732
Principal Component 3 (PC3)	7.47957	0.05123
Principal Component 4 (PC4)	6.37829	0.04369
Principal Component 5 (PC5)	5.9081	0.04047
Principal Component 6 (PC6)	5.45423	0.03736
Principal Component 7 (PC7)	5.16522	0.03538
Principal Component 8 (PC8)	3.76667	0.0258
Principal Component 9 (PC9)	3.6941	0.0253

TABLE II
SCORES FOR PC1 AND PC2 (COMPONENT 1 & 2) OF CU01

Sample Number	Scores for PC1	Scores for PC2	Class
1	-1.18113	-2.54187	abnormal
2	-1.26634	-2.64893	abnormal
3	-0.88636	-2.97021	abnormal
4	-0.12991	-1.99849	abnormal
5	3.246313	1.579908	abnormal
6	2.38746	0.798551	abnormal
7	-4.92702	0.419675	normal
8	-5.10988	-0.03882	normal
9	-4.82858	0.275417	normal
10	-4.99787	-0.12245	normal
11	-5.02603	0.221886	normal
12	-4.94025	0.02097	normal

Using the procedure discussed earlier the Table II for a ventricular tachycardia patient (CU01) that shows the first two principal component scores for every normal and abnormal ECG segment. Since 6 normal and another 6 abnormal tests

are performed ECG segments for every patient, 12 sets of values for the principal component scores are obtained. This particular Table II corresponds to patient CU01 of CU Ventricular Tachyarrhythmia Database. Similarly, scores can be derived for other patients in the database. It is obvious from this distribution that abnormal ECGs can be easily separated from the normal ones. Similar tests were performed on all other patients from Ventricular Tachyarrhythmia Database, and they all follow the same trend which confirms that abnormal ECGs can be distinguished from the normal ECGs when PCA is applied to compressed ECGs of the patients.

TABLE III
K-MEAN RESULTS FOR CU01

Sample Number	Distance from class1	Distance from class2	Class
1	21.50359	3.930326	2 (abnormal)
2	21.44839	4.478564	2 (abnormal)
3	26.29707	4.35777	2 (abnormal)
4	27.9701	0.733969	2 (abnormal)
5	69.63799	16.59684	2 (abnormal)
6	54.60353	8.494487	2 (abnormal)
7	0.086221	30.91665	1 (normal)
8	0.047432	31.52044	1 (normal)
9	0.041762	29.41073	1 (normal)
10	0.064142	30.10386	1 (normal)
11	0.011507	31.33386	1 (normal)
12	0.01275	29.84699	1 (normal)

Table III shows the results for k-mean algorithm as it is applied to previous data shown in Table II. From the results it is clear that the distances of samples 1-6 are small for class 2 (abnormal) and large for class 1 (normal). This is why it is classified as class 2 (abnormal). Similarly, samples 7-12 have small distance from class 1 and large distance from class 2. Therefore, it is classified as class1 (normal).

V. RESULTS AND CONCLUSION

In k_mean algorithm depending on the distances from class1 and class2 values of various samples. The distances are small for abnormal signals and large for normal signals. The benefit of diagnosis from compressed ECG is immense. As compressed ECG contains less characters, diagnosis from compressed ECG can be possible (using the techniques shown in [18]) Most importantly, for telecardiology applications, where ECG is transmitted and stored in compressed format, cardiovascular diagnosis is possible, without performing decompression, saving processing power, resource and time. Minimizing delays in diagnosis entail savings of patients' lives.

REFERENCES

- [1] G. Clifford, F. Azuaje, and P. McSharry, Advanced methods and tools for ECG data analysis. Artech House.
- [2] data compression for storing and transmitting ECG/VCGs ' by M.E.Womble and J.S.Halliday
- [3] "application of a partitioned karhunen-loeve expansion scheme to ECG/VCG data compression " by A.M.ZIED and E.Womble
- [4] E. Ubeyli, "Eigenvector Methods for Automated Detection of Electrocardiographic Changes in Partial Epileptic Patients
- [5] W. Jiang and S. Kong, "Block-based neural networks for personalized ECG signal classification," *IEEE Transactions on Neural Networks*
- [6] Cardiac disorder diagnosis based on ECG segments analysis & classification" by Sheik.R.R.&Taj.I.A
- [7] ECG waveform data extraction from paper ECG recordings by k-means method" by Guojie Shi; Gang Zheng; Min Dia;
- [8] Two novel methods for multiclass ECG arrhythmias classification based on PCA, fuzzy support vector machine & unbalanced clustering" by Nait-Hamoud.M.C. & Moussaoui.A.