PERFORMANCE ANALYSIS OF FEATURE EXTRACTION SCHEMES FOR ECG SIGNAL CLASSIFICATION

DEVASHREE JOSHI, RAJESH GHONGADE

Department of Electronics & Telecommunication Engineering, Vishwakarma Institute of Information Technology, Pune, (Maharashtra), India.
Email: joshi.devashree@yahoo.in, rbghongade@gmail.com

Abstract—Electrocardiogram (ECG) is the P, QRS, T wave indicating the electrical activity of the heart. Electrocardiogram is the most easily accessible bioelectric signal that provides the doctors with reasonably accurate data regarding the patient heart condition. Many of the cardiac problems are visible as distortions in the electrocardiogram (ECG). Normally ECG related diagnoses are carried out manually. As the abnormal heart beats can occur randomly it becomes very tedious and time-consuming to analyze say a 24 hour ECG signal, as it may contain hundreds of thousands of heart beats.

In this work we propose computer based automated system to help the doctor to detect cardiac arrhythmia. As reference, we have used the Normal, Premature Ventricular Contraction (PVC) and Fusion signals of the MIT-BIH Database. Then we have focused on the various schemes for extracting the useful features of the ECG signals for use with artificial neural networks. We extract the principal characteristics of the signal by means of the Principal Component Analysis (PCA) technique and other techniques such as Discrete Wavelet Transform and Discrete Cosine Transform. After signal pre-processing, they are applied to an Artificial Neural Network Multilayer Perceptron (ANN MLP). The task of an ANN based system is to correctly identify the three classes the feature extraction schemes are discussed and compared with RBFN & Support Vector Machine in this work.

Keywords—ECG, MLP, Feature extraction, DWT, DCT, PCA, SVM

I. INTRODUCTION

Electrocardiography deals with the electrical activity of the heart. Monitored by placing sensors at limb extremities of the subject, the electrocardiogram (ECG) is a record of the origin and propagation of electrical potential through cardiac muscles. It is considered a representative signal of cardiac physiology, useful in diagnosing cardiac disorders. The biggest challenge faced by the models for automatic heart beat classification is the variability of the ECG waveforms from one patient to another even within the same person. However, different types of arrhythmias have certain characteristics which are common among all the patients. Thus the objective of a heart beat classifier is to identify those characteristics so that the diagnosis can be general and as reliable as possible. One of such methods which can be reliably used for ECG classification is the use of neural networks. [2]

Figure 1 depicts the three types of heartbeats. Here we discuss the various schemes to extract useful information from the ECG signal and use these features for training an ANN for pattern recognition of these three types of heartbeats and classify them correctly as Normal (N), Fusion (F) and Premature Ventricular Contraction (PVC). We have used the ECG data available from MIT/BIH Arrhythmia Database since it is considered as the benchmark data. [11]

II. METHODOLOGY

Fig. 3 depicts the proposed methodology of classification of Normal and two types abnormal beats in the ECG of arrhythmia. The working of each block is explained in the following sections.
A. ECG DATA PRE-PROCESSING

The ECG signal downloaded from MIT-BIH arrhythmia database [11] may contain artefacts, noise and baseline wanderers. Therefore it is necessary to denoise the ECG signal to remove all these unwanted parts of the signal. After denoising the ECG, it is subjected to QRS complex detection using Pan Tompkins algorithm. The QRS complex is physiologically an important peak in the ECG signal, also it is easy to detect by signal processing algorithms due to its sharp and prominent shape. In this study we have used Pan Tompkins algorithm for detection of QRS complex (Pan & Tompkins, 1985). [10] The algorithm consists of computation of derivatives, moving window integrator, squaring and detection of rising edge of pulses. The derivative provides the slope information of ECG waveform, squaring will emphasize higher amplitudes and suppresses smaller amplitudes and moving window integrator performs averaging operation, thereby removes noise.

After detection of QRS complex, 90 samples were chosen from the left side of QRS mid-point and 90 samples after QRS mid-point. We can use the entire 180 samples for training an ANN but this may give rise to two problems namely overtraining and excessive computational overhead. Overtraining affects the accuracy adversely while computational overhead will put constraints on the speed and required resources. Hence only useful features have to be identified so that only these features can be used to train an ANN and avoid the above mentioned problems. [5]

![Figure 3. Block Diagram of proposed system.](image)

B. PRINCIPAL COMPONENT ANALYSIS

PCA is also known as Karhunen-Loève Transform - KLT. The Principal Component Analysis (PCA) is a linear dimensionality reduction technique that provides projection of the data in the directions of highest variance. The importance of the variables is statistically evaluated. So, this method is based on the significance of the information, in which the KLT identifies the direction of the signal with maximum energy or variance. Here again the entire data set is subjected to principal component analysis. We get 180 components corresponding to the dimensionality of the input sequence. Components those are significant from the point of view of contribution to the total energy of the signal are selected. The selected components together must contribute about 99% of the total energy of the signal. This procedure decreases the data dimensionality without significant loss of information. [4]

PCA ALGORITHM:

a. Obtain the Mean vector and mean adjusted data:
\[
\mu_x = \frac{1}{N} \sum_{i=0}^{N-1} x_i
\]

Mean Adjusted Data = \( X - \mu \)

b) Obtain the covariance matrix:
\[
C_x = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \mu_i)(x_i - \mu_i)^T
\]

c) Obtain the Eigen vectors and Eigen values
\[
C_x e = \lambda
\]

Where e is Eigen vector and \( \lambda \) is Eigen value.

d) Choosing components and forming a feature vector:

We get 180 components corresponding to the dimensionality of the input sequence. Components that are significant from the point of view of contribution to the total energy of the signal are selected. The selected components together must constitute about 99% of the total energy of the signal. This procedure decreases the data dimensionality without significant loss of information.

\[
\text{Feature Vector} = (eig_1 \ eig_2 \ eig_3 \ldots eig_P)
\]

e) Creating the new data set:
\[
\text{New Data} = [\text{Feature Vector}] T \times \text{Mean Adjusted Data}
\]

Once we select 14 significant components the next task is to train a MLP with these coefficients as inputs and the category of the heartbeat as output.

C. DISCRETE COSINE TRANSFORM

The discrete cosine transform is applied to the individual ECG beats. The frequency coefficients are estimated as follows:
\[
Y[u] = G[u] \sum_{i=0}^{N-1} y[i] \frac{\cos[(2i+1)u]}{2N}
\]

Where \( N \) is the length of the signal \( y[i] \) for \( i = 0, 1 \ldots (N-1) \). For the AC/DCT method \( y[i] \) is the auto correlated ECG. \( G[k] \) is given from:
With the energy compaction property of DCT allows representation in lower dimensions. This way, near zero components of the frequency representation can be discarded and the number of important coefficients is eventually reduced.

DCT offers a powerful feature extraction mechanism. The individual heartbeats are subjected to 1-dimensional DCT. Here only 30 DCT coefficients were selected again based on the morphology preservation with a metric of PRD and the energy content.[4]

$$PRD = \frac{\sqrt{\sum_n (\hat{x}_n - x_n)^2}}{\sqrt{\sum_n x_n^2}}$$  \hspace{1cm} (3)

Where, $\hat{x}$ is reconstructed ECG signal and $x$ is original ECG signal.

D. Discrete Wavelet Transform
The DWT analyses the signal at different resolution (hence, multiresolution) through the decomposition of the signal into several successive frequency bands. The DWT utilizes two set of functions $\Phi(t)$ and $\Psi(t)$, each associated with the low pass and the high pass filters respectively [8]. These functions have a property that they can be obtained as the weighted sum of the scaled (dilated) and shifted version of the scaling function itself:

$$\varphi(t) = \sum_n h[n]\varphi(2t - n)$$  \hspace{1cm} (1)

$$\Psi(t) = \sum_n g[n]\varphi(2t - n)$$  \hspace{1cm} (2)

Here, $h[n]$ and $g[n]$ is the half band low pass filter and high pass filter respectively.

Wavelet analysis consists of decomposing a signal or an image into a hierarchical set of approximations and details. Daubeches D4 wavelet is chosen as the mother wavelet since it resembles a heartbeat waveform. Again 23 coefficients are selected on the basis of morphology preservation criterion.[4]

Once these 23 DWT coefficients are selected the next task is to train a MLP with these coefficients as inputs and the category of the heartbeat as output.

III. Artificial Neural Networks
An artificial neural network is inspired by the biological neurons present in the human brain. Though the sheer number of biological neurons and their high interconnectivity is impossible to be duplicated, the scaled down models of the artificial neural networks shows similar capacity of learning and generalizing intrinsic characteristics of the information presented to it.[4]

A. Multilayer Perceptron (MLP)
This is one of the most popular and powerful ANN architectures. MLP not only overcomes the severe limitation of a simple perceptron, but also offers the most important feature of learning. It is thus a very powerful pattern classifier. Function approximation can also be done quite effectively by the MLP.[7]

Figure 4 depicts the MLP architecture used for experimentation.

The entire process of classification consists of two phases: training phase and testing phase. MLP training is executed using a well-known technique of back propagation. After satisfactory training error level is obtained the training is terminated and a new data is presented to the network to test the performance.

a. BACK PROPAGATION ALGORITHM
Following steps indicate the back propagation algorithm. [1]

a) Initialize the weights and biases with small random values.

b) Present a continuous valued input vector $X_0$, $X_1$,...,$X_{n-1}$ and specify the desired outputs $d_0, d_1$,...,$d_M$.

c) Compute actual outputs $Y_0, Y_1$,...,$Y_{M-1}$.

d) Adapt weights: Use a recursive algorithm starting at the output nodes and working back to the first hidden layer. Adjust weights by

$$W_{ij}/(t+1) = W_{ij}(t) + \eta \delta_j X_i'$$  \hspace{1cm} (1)

Where $W_{ij}(t)$ is the weight from hidden node $i$ or from an input to node $j$ at time $t$, $X_i'$ is either the output of node $i$ or is an input, $\eta$ is a gain term and $\delta_j$ is an error term for node $j$.

If node $j$ is an output node then

$$\delta_j = (1 - (Y_j)^2) d_j$$  \hspace{1cm} (2)

Where $d_j$ is the desired output of node $j$ and $Y_j$ is the actual output.

e) If node $j$ is an internal hidden node then

$$\delta_j = (1 - (X_j)^2) \sum_k \delta_k W_{j,k}$$  \hspace{1cm} (3)

where $k$ is over all the nodes in the layers above node $j$. 

Performance Analysis Of Feature Extraction Schemes For ECG Signal Classification
f) Internal node thresholds are adapted in a similar manner by assuming they are connection weights on links from auxiliary constant valued inputs (called as biases). Convergence is faster if a momentum term is added and weight changes are smoothed by

\[ W_{ij}(t+1) = W_{ij}(t) + \eta \Delta I_i + \alpha [W_{ij}(t) - W_{ij}(t-1)] \quad (4) \]

Where \(0 < \alpha < 1\)

Here we have used log sigmoid functions at the output of each neuron. It is given by the following equation:

\[ y = \frac{1}{1+e^{-x}} \]

If the mean square error (MSE) is not acceptable then, go back to step b.

**B. RBF NETWORKS**

Radial basis neural networks were evaluated for deriving the optimum network performance. A variation in the cluster sizes was carried out and the accuracy was found to be best for 150 clusters. The comparative performance of the three RBF based classifiers is presented in Table II.

**C. SUPPORT VECTOR MACHINE**

This paper focuses on implementing a new arrhythmia classification algorithm which will be able to effectively identify two different classes of cardiac arrhythmia. Support Vector Machine (SVM) performs classification by constructing a multi dimensional hyper plane that optimally separates the data into two categories. Using a kernel function, SVM’s are an alternative training method for polynomial, radial basis function and multi-layer perceptron classifiers in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem. SVM is a machine-learning technique which has established itself as a powerful tool in many classification problems. Simply stated, the SVM identifies the best separating hyper plane (the plane with maximum margins) between the two classes of the training samples within the feature space by focusing on the training cases placed at the edge of the class descriptors. In this way, not only an optimal hyper plane is fitted, but also less training samples are effectively used; thus high classification accuracy is achieved with small training sets .[5]

Given a training set

\[(X_i, Y_i), i = 1, 2, ..., l \text{ where } X_i \in \mathbb{R}^n \text{ & } Y_i \in (-1,1)\]

traditional SVM algorithm is summarized as the following optimization problem:

\[
\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{l} \xi_i \quad y_i f(x_i) \geq 1 - \xi_i, \quad (1)
\]

for all \(i \quad \xi_i \geq 0\)

SVM classification, Dual formulation:

\[
\min_{\alpha_i} \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j)
\]

\[0 \leq \alpha_i \leq C, \text{ for all } i; \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \quad (2)\]

Variables \(\xi_i\) are called slack variables and they measure the error made at point \((x_i, y_i)\).

The decision function can be written by

\[f(x) = \text{sign}(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x_i)) + b \quad (4)\]

Training SVM becomes quite challenging when the number of training points is large. A number of methods for fast SVM training have been proposed. For training, we are using SMO support vector machine method. This formulation leads to an extremely fast and simple algorithm for generating linear or nonlinear classifier. Results are shown in table II.

**IV. EXPERIMENTATION AND RESULTS**

We used the transformed data obtained using the above mentioned techniques to train a MLP. The MLP architecture consisted of one hidden layer with 15 neurons in the hidden layer and three output neurons corresponding to the three classes under consideration. However the input neurons corresponding to the four methods were different since number of features so obtained varied with the extraction scheme. The training set consisted of equal samples of each type to reflect the equal probability data. Multifold and differential training was employed to enable true network learning and remove the dataset bias if any. Rigorous experimentation indicated that momentum learning rule for MLP and tan-sigmoid activation function produced best results. Optimum number of hidden layer neurons and training iterations was explored experimentally. Naturally the four feature extraction techniques exhibited different optimum configurations. Memorizing due to overtraining was avoided by again subjecting the networks to iterations ranging from 1000 to 3000 and finding the iteration number after which the testing accuracy deteriorated. Presented in Fig.5 Once the training phase completed we tested the trained network with 600 samples (containing 200 samples of each type). Thus we subjected the network to equal sample test data. The comparative performance of the three ANN based classifiers is presented in Table I.
TABLE I
PERFORMANCE OF ANN CLASSIFIER

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PCA</td>
<td>100</td>
<td>99.05</td>
<td>97</td>
<td>98.83</td>
</tr>
<tr>
<td>2</td>
<td>DCT</td>
<td>97.50</td>
<td>99.50</td>
<td>97.50</td>
<td>98.16</td>
</tr>
<tr>
<td>3</td>
<td>DWT</td>
<td>100</td>
<td>98.5</td>
<td>97</td>
<td>98.5</td>
</tr>
</tbody>
</table>

Statistical measures of the performance of a binary classification can be measured using Sensitivity, Specificity, PP (Positive Predictivity) and CR (Classification rate). Presented in Fig. 6. Sensitivity is the fraction of abnormal ECG beats that are correctly detected among all the abnormal ECG beats. Specificity is the fraction of normal ECG beats that are correctly classified among all the normal ECG beats. Specificity measures the proportion of negatives which are correctly identified. Positive Predictivity (PP) is the fraction of real abnormal ECG beats in all detected beats. The specificity of the test is equal to 1 minus the false positive rate. Classification rate (CR) is the fraction of all correctly classified ECG beats, regardless of normal or abnormal among all the ECG beats.

TABLE II
PERFORMANCE OF RBF & SVM CLASSIFIER

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Method</th>
<th>Number of Neurons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Input</td>
</tr>
<tr>
<td>1</td>
<td>PCA</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>DCT</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>DWT</td>
<td>23</td>
</tr>
</tbody>
</table>

DISCUSSIONS

ANN based classifiers are now an established paradigm for pattern classifiers, MLP being a simple but powerful architecture. The main power and robustness of MLP can be harnessed by proper selection of input features. All the ECG sample values per heartbeat can be presented to MLP but this consumes a lot of computational resources and training time. This may also over train network and prevent it from generalizing. Feature extraction proves an essential process for reducing the inputs to the ANN drastically. Also the bare minimum and prominent markers are obtained which enhance the performance of the ANN classifier in addition to making it more robust and fault tolerant. The MLP generalizes properly as can be seen from the results.

Figure 5 Graphical representation of Accuracy vs. Iterations.

Figure 6. Classifier performance in terms of PP, specificity, CR and Sensitivity for DCT.

The performance parameters based on confusion matrix were also calculated for DCT and plotted as indicated by the Figure 6. Sensitivity and specificity of Normal type is higher than other two. Positive predictivity for PVC is higher and classification rate of PVC is higher. Similarly we can find statistical measurements of performance parameters using

TABLE III RESOURCES CONSUMED IN TERMS OF NEURONS FOR ANN

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Method</th>
<th>% Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>PVC</td>
</tr>
<tr>
<td>1</td>
<td>PCA + RBF</td>
<td>95.23</td>
</tr>
<tr>
<td>2</td>
<td>DCT + RBF</td>
<td>78.23</td>
</tr>
<tr>
<td>3</td>
<td>DWT + RBF</td>
<td>82.13</td>
</tr>
<tr>
<td>4</td>
<td>PCA + SVM</td>
<td>99.50</td>
</tr>
<tr>
<td>5</td>
<td>DCT + SVM</td>
<td>99.50</td>
</tr>
<tr>
<td>6</td>
<td>DWT + SVM</td>
<td>99.00</td>
</tr>
</tbody>
</table>
feature vectors of methods like DWT and PCA as shown in figure 7 and 8. Classifier performance for DWT shows that classification rate of Normal type is higher and specificity is highest. If we compare classifier performance of PCA with other two we see that specificity and positive predictivity of PCA is highest for all pattern types. So PCA gives good results.

Classifier performance for DWT shows that classification rate of Normal type is higher and specificity is highest. If we compare classifier performance of PCA with other two we see that specificity and positive predictivity of PCA is highest for all pattern types. So PCA gives good results.

Figure 7. Classifier performance in terms of PP, specificity, CR and Sensitivity for DWT.

Figure 8. Classifier performance in terms of PP, specificity, CR and Sensitivity for PCA.

NOISE ANALYSIS

As an additional metric we subjected the classifier to noise analysis, wherein the performance of the classifiers were evaluated by contaminating the original ECG signal with white Gaussian noise of varying strength. The noise performance is depicted in figure 9.

CONCLUSION

SVM classifiers are definitely superior in performance and robustness than the traditional classifiers like Multilayer perceptron. A definite improvement in the accuracy can be achieved by selecting the optimum network topology, learning rule, number of iterations for training and learning rate. Another important finding is that the feature extraction techniques play a crucial role in the performance of these classifiers. Discrete cosine transform offers a robust and near optimal method for selecting features of the ECG signal. MLP with DCT and PCA is a promising combination for a robust ECG signal classifier. Future work will focus on algorithm for speed optimization and further optimization of the feature extraction and classification scheme.

REFERENCES

[1] Qu Xiao; Cai Wei Jian; Ge Ding Fei, “ECG signal classification based on BPNN”, IEEE Conference Publications, 2011 International Conference on Digital Object Identifier: ICEICE.


Performance Analysis Of Feature Extraction Schemes For ECG Signal Classification

---


