Classification of ECG signal during Atrial Fibrillation using Autoregressive modeling

K.Padmavathi\textsuperscript{a}, K.Sri Ramakrishna\textsuperscript{b}

\textsuperscript{a}Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad 500090, India
\textsuperscript{b}VR Siddhardha Engineering College, Vijayawada 520007, India

Abstract

Atrial fibrillation (AF) is a common type of arrhythmia that causes death in the adults. The auto regressive (AR) coefficients characterize the features of AF. The AR coefficients are measured for every 15 second duration of the ECG and the features are extracted using Burg’s method. These features are classified using the different statistical classifiers such as kernel Support Vector Machine (KSVM) and K-Nearest Neighbor (KNN). The performance of these classifiers is evaluated on signals obtained from MIT-BIH Atrial Fibrillation Database. The effect of AR model order and data length is tested on the classification results.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: Atrial Fibrillation; AR coefficients; Burg method; SVM; KNN; MIT/BIH database.

1. Introduction

Computerized electrocardiogram classification can help to reduce healthcare costs. ECG results indicate the presence of AF alarming the status of patient’s heart. During AF, the hearts atria are quicker than normal beating. As the blood is not ejected completely out of atria, there might be chances of formation of blood clots in the atria resulting in increased risk of stroke. Electrocardiogram (ECG) is one of the useful tool for AF detection. AF can be detected by observing three main morphological features in the ECG as shown in Fig.1. They are

- P wave absence.
- Instead of P waves fluctuating waveforms (f-waves).
- Heart rate irregularity.
There are several methods to detect the features of AF\(^9\). Methods based on RR interval are proposed in\(^1,2\). P wave based methods are presented in\(^1,25\). The RR interval, P wave based methods have some limitations\(^8\). When the ECG changes quickly between rhythms or when Atrial Fibrillation takes place with regular ventricular rates, the methods based on RR interval fail in accurate detection\(^2\). Detecting the absence of P wave is difficult due to its small amplitude\(^25\). To study the atrial activity during AF, frequency domain techniques have been proposed in\(^22,19,21,20\). Ventricular activity needs to be canceled before applying FFT. In presence of noise\(^20\) this cancellation process may be difficult and involves high computation. Morphological features are difficult to detect because they change from patient to patient. From statistical features (AR features) we can easily classify AF signals. AR coefficients\(^24\) are the simplest and best features for AF classification. This paper emphasizes on the use of AR modeling to discriminate between Non-AF and AF waves. Previous studies claim that, the usage of AR coefficient features yield better results than original time series features\(^24,23\). The proposed algorithm is estimated based on the data segments collected from MIT-BIH Atrial Fibrillation Database\(^16\). The AF classification flow diagram as shown in Fig.2.

2. Preprocessing

2.1. Data

In order to assess the performance of the algorithm, Physionet Atrial Fibrillation Database\(^16\) is used, which consists of 23 AF recordings at a sampling rate of 250Hz and 18 Normal Sinus rhythm recordings at 128Hz sampling rate. Before feature extraction AF signals are resampled at 128Hz. The AF signal which is at 250Hz is resampled at 128Hz as shown in Table 1.
2.2. Noise Removal

The first step in this algorithm is dividing the signal into desired length. After segmentation, each segment is considered column matrix for compact notation and baseline wander present in the signal is removed with the help of sgolay filtering\(^\text{17}\) as in Fig. 3.

3. Feature Extraction

3.1. Computation of AR coefficients

Autoregressive model is based on the principle of linear prediction. In AR model\(^\text{17}\) each sample is predicted based on the linear combination of previous samples. Let \(f_1, f_2, f_3, \ldots, f_n\) be the time series. The \(p\)th order autoregressive time series (written as AR\((p)\)) of \(F(n)\) is given by the the equation.

\[
F(n) = \sum_{j=1}^{p} \alpha_j F(n-j) + \epsilon(n) \tag{1}
\]

Where \(P\) is the model order \(\epsilon(n)\) is assumed to be white Gaussian noise with zero mean and variance \(\sigma^2\). The AR model parameters \(\alpha_j\) are calculated using Yule-Walker, Burgs methods and the selected model order experimentally.

3.2. Yule-walker Method (YW)

\[
\epsilon(n) = F(n) - \hat{F}(n) \tag{2}
\]

\(\hat{F}(n)\) is Predicted value \(F(n)\)

\[
E = \sum_{n=1}^{N} (\epsilon(n))^2 = \sum_{n=1}^{N} (F(n) - \hat{F}(n))^2 \tag{3}
\]
\[ E = \sum_{n=1}^{N} (F(n) - \sum_{n=1}^{N} \alpha_j F(n - j))^2 \]  

(4)

\( \alpha_j \) is predicted to minimize error \( \epsilon(n) \). Mean square value of the error will be minimum if \( \frac{\partial E}{\partial \alpha_j} = 0 \)

\[ \sum_{j=1}^{p} \alpha_j R(j - i) = R(i) \]  

(5)

\[ R\alpha = r \]  

(6)

\[ \alpha = R^{-1}r \]  

(7)

3.3. Burg’s method

Input signal \( F(n), n=1,2,\ldots,N \), and let us consider the backward and forward linear predictions of order \( k=1,2,\ldots,m \)

\[ \hat{F}(n) = -\sum_{k=1}^{m} \alpha_m(k) F(n - k) \]  

(8)

\[ \hat{F}(n - m) = -\sum_{k=1}^{m} \beta_m(k) F(n - m + k) \]  

(9)

where \( \alpha_m \) and \( \beta_m \) are the forward and backward prediction coefficients respectively

\( f^t(n) = [f(n), f(n-1), \ldots, f(n-p)] \).

\[ f_m(n) = F(n) - \hat{F}(n) = \sum_{k=1}^{m} \alpha_m(k) F(n - m) \]  

(10)

\[ b_m(n) = F(n - m) - \hat{F}(n - m) = \sum_{k=1}^{m} \beta_m(k) F(n - m + k) \]  

(11)

Where \( \alpha_m \) and \( \beta_m \) are the forward and backward prediction residuals. Note that \( \alpha_m = 1, \beta_m = 1 \) by definition. The FIR prediction error filter or the lattice filter is given by the set of recursive equations

\[ f_m(n) = f_{m-1}n + k_m b_{m-1}(n - 1) \]  

(12)

\[ b_m(n) = k_m f_{m-1}(n - 1) \]  

(13)

\( m=1,2,3,\ldots,p \). Where \( K_m \) are the reflection coefficients of the \( m^{th} \) recursion step. The initial values of the residuals are

\( f_0(n) = b_0(n) = f(n) \)

\[ k_m = \frac{-2 \sum_{n=p+1}^{N} [f_{m-1}(n) + k_m b_{m-1}(n - 1)]}{\sum_{n=p+1}^{N} [(f_{m-1}(n))^2 + b_{m-1}(n)^2]} \]  

(14)

\[ \alpha_m(k) = \alpha_{m-1}(k) + k_m \alpha_{m-1}(k - m) \]  

(15)

\( \alpha_m(0) = 1, \alpha_m(m) = k_m \),

where \( m=1 \) to \( p \) and \( k=1 \) to \( m \).

All-pole prediction coefficients method excel in comparison to the autocorrelation method because they decrease the total prediction errors and the data sequence is not subjected to any window function. The advantage of the former method is that it is computationally efficient, stable and has high frequency resolution. The selection of the Autoregressive model order is of foremost importance in the classification of AF. The correct number of Autoregressive coefficients are determined using trial and error method. The coefficients of order 4,8,16 are used for our study.
4. Classification

The performance of two different classifiers SVM and k-NN are obtained with the AR coefficients as input.

4.1. Kernel Support Vector Machines (KSVM)

A kernel Support Vector Machine is a supervised machine learning technique applicable for classification. It is an example for non-probabilistic binary linear classifier, established from Statistical Learning Theory. It exhibits high accuracy and has capability to deal with high dimensional data sequences. The support vector machine makes use of pattern recognition among two point classes by Support Vectors (SV).

Kernels are functions that perform some mathematical operations on \( x_1, x_2 \) depending on the selection of the kernel function.

\[
    x = \phi(x) \quad (16)
\]

\[
    k(x_1, x_2) = \phi(x_1)\phi(x_2) \quad (17)
\]

Gaussian kernel can be expressed as

\[
    k(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1 - x_2)^2}{2\sigma^2}} \quad (18)
\]

Linear kernel can be expressed as

\[
    K(x_1, x_2) = \phi(x_1)\phi(x_2) \quad (19)
\]

\[
    K(x_1, x_2) = x_1'x_2 \quad (20)
\]

Kernel functions can be applied to non-linear data so that non-linear features are converted into linear features as shown in Fig. 4. By using kernel trick features can be represented in a high dimensional feature space. Linear classifier methods used to produce non-linear classification is the major advantage of kernels. As ECG is an one dimensional signal, \( x_1 \) is \( x_2 \) are the features of a two distinct ECG recordings. Linear and Gaussian kernels are applied on the ECG signal with SVM classifier.

4.2. K-Nearest Neighbour (KNN)

In the K-nearest neighbors rule, a new vector \( y \) of a new class is classified based on the distance from nearest mean vector. The distance from vector \( y \) and the centroid of the \( m^{th} \) cluster \( z_l \) is calculated as the Euclidean distance

\[
    s_m = \sqrt{\sum_{l=1}^{n} (y_l - z_l^m)^2} \quad (21)
\]
Table 1. MIT-BIH Record Numbers

<table>
<thead>
<tr>
<th>Normal Data</th>
<th>AF Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>16265,16272,16273,</td>
<td>04015,04043,04048,</td>
</tr>
<tr>
<td>16420,16483,16539,</td>
<td>04126,04746,04908,</td>
</tr>
<tr>
<td>16773,16786,16795,</td>
<td>04936,05091,05121,</td>
</tr>
<tr>
<td>17052,17453,18177,</td>
<td>05261,04426,06453,</td>
</tr>
<tr>
<td>18184,19088,19090,</td>
<td>06995,07162,07859,</td>
</tr>
<tr>
<td>19103,19140,19830.</td>
<td>0787,07910,08215,</td>
</tr>
<tr>
<td></td>
<td>08219,0837,08405,</td>
</tr>
<tr>
<td></td>
<td>08434,08455.</td>
</tr>
</tbody>
</table>

Table 2. Classification for Model order 4

<table>
<thead>
<tr>
<th>Data length</th>
<th>YW+SVM</th>
<th>Burg+SVM</th>
<th>YW+KNN</th>
<th>Burg+KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Sec</td>
<td>76.9</td>
<td>92.3</td>
<td>46.1</td>
<td>38.4</td>
</tr>
<tr>
<td>15 Sec</td>
<td>92.3</td>
<td>76.9</td>
<td>92.3</td>
<td>92.3</td>
</tr>
<tr>
<td>30 sec</td>
<td>92.3</td>
<td>92.3</td>
<td>76.9</td>
<td>76.9</td>
</tr>
</tbody>
</table>

Table 3. Classification for Model order 6

<table>
<thead>
<tr>
<th>Data length</th>
<th>YW+SVM</th>
<th>Burg+SVM</th>
<th>YW+KNN</th>
<th>Burg+KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Sec</td>
<td>76.9</td>
<td>92.3</td>
<td>53.8</td>
<td>61.5</td>
</tr>
<tr>
<td>15 Sec</td>
<td>84.6</td>
<td>100</td>
<td>69.2</td>
<td>92.3</td>
</tr>
<tr>
<td>30 sec</td>
<td>92.3</td>
<td>100</td>
<td>100</td>
<td>92.3</td>
</tr>
</tbody>
</table>

Table 4. Classification for Model order 8

<table>
<thead>
<tr>
<th>Data length</th>
<th>YW+SVM</th>
<th>Burg+SVM</th>
<th>YW+KNN</th>
<th>Burg+KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Sec</td>
<td>76.9</td>
<td>92.3</td>
<td>84.6</td>
<td>100</td>
</tr>
<tr>
<td>15 Sec</td>
<td>84.6</td>
<td>100</td>
<td>84.6</td>
<td>100</td>
</tr>
<tr>
<td>30 sec</td>
<td>92.3</td>
<td>100</td>
<td>92.3</td>
<td>100</td>
</tr>
</tbody>
</table>

m is the cluster index, n is the number of the parameters used and l the parameter index. Vector y can be classified in to class k at which $s_m$ is minimum. We selected the value of k as 1.

5. Results

The 15,30 and 60 second length sequences from each recording are considered and AR coefficients are calculated. The effect of model order on classification results is investigated. For the SVM and K-NN classifiers, 280 recordings are given for training (2/3 of total recordings) and 93 (1/3 of total recordings) are given for testing. Three modeling orders are used to differentiate the proposed method with other methods. Classification accuracies for different model orders and different lengths are shown in 2 to 4 Tables.

The results of these methods are shown in Tables 2 to 4. It is evident that the burg’s method with KNN classifier shows best results among the two classifiers irrespective of the length of data sequence for model order 8.

6. Conclusion

In this paper the use of AR modeling for Atrial Fibrillation arrhythmia detection is examined. A comparison of the performance of SVM and kNN classifiers on signals from MIT-BIH Atrial Fibrillation Database is depicted. Analysis effect of various model order’s for different data segment lengths is performed. Among the two classifiers KNN with Burg’s method achieved the best results. The minimum misclassified segments were achieved in 5,15,30 second
segments for the model order 8, which proves to be the best classification obtained. Burg’s method shows good results for short data segments with SVM classifier. Yule Walker method shows good results for data segments of length 30 seconds for model order 6 with KNN classifier. Selecting the model order and segment length depends on the required precision and availability of the computational resources. This algorithm can be used for real time detection of AF signals. The former procedures for feature extraction such as ventricular activity cancellation and detection of R peak, which are tedious in nature can be eliminated.

References