Classification of Normal and PVC-heart beat Using DCT-Time Domain Features and SVM classifier

Abdorasoul Ghasemi[1], Karim Faez[1]
[1]Electrical Engineering Department, AmirKabir University of Technology, Hafez Avenue, Tehran, Iran, 15914
{arghasemi, kfaez}@ aut.ac.ir

Abstract: *This paper describes a method for classification of normal (N) and premature ventricular contraction (PVC) heart beats based on SVM classifier. We Construct the feature vector such that it contains three types of information: the shape or morph information, the intra beat time domain information and the inter beat time domain variations. We use DCT coefficients of the ECG segments as the shape information, the linear predictive modeling to represent the intra beat time dynamics of each segment and RR intervals to consider the inter beats time domain variations. The proposed method to construct the feature vector in conjunction with the SVM classifier, that has a good generalization property, is used to classify the Normal and PVC heart beats. Evaluations are done using the MIT BIH arrhythmia database and the achieved classification accuracy is 98.5% for testing data.*

Keywords: ECG signal, DCT coefficient, linear predictive modeling, SVM classifier

1. INTRODUCTION

The Electrocardiogram (ECG) is the electrical activity of the heart that can be recorded for diagnosing heart diseases. Physicians diagnose heart diseases based on the ECG signal shape and distances between fiducial points to decide whether the heart beat belong to the normal, or to the appropriate class of arrhythmia. Automatic recognition and interpretation of ECG signals was widely investigated by many researchers to help reduce the burden of physicians. Typical cycle of heat normal beat is shown in Fig. 1.a. Fig. 1.b shows a specific heart beat arrhythmia (PVC beat) that we aim to distinguish them from the neighboring normal beats.

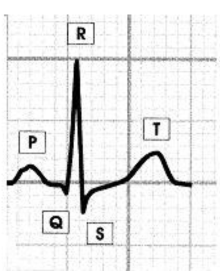

Figure 1.a: fiducial points on the normal beat

Many algorithms have been proposed over past years to develop automatic systems for accurately classify the ECG signals. They differ on the type of the applied signal processing techniques to extract discriminative features of the signal and the classification method. Various methods for segmentation and feature extraction have been proposed that use Time, Frequency or both domain information such as STFT [1], wavelet transform [2] and Hermit basis functions expansion [3].

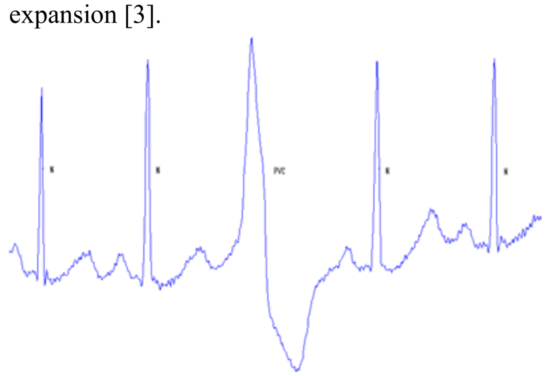

Figure 1.b: PVC beat between normal beats

Due to high variations in the signals among the same types of beats, it is difficult to separate one from other based on only time or frequency representation.

In addition, various classifiers mainly based on neural solutions were developed [4] to classify the extracted features, the best known of them include multilayer perceptron and kohonen self organizing network. Recently SVM classifier has been widely used due to its good generalization performance in different pattern recognition problems.

DCT features are widely used for compression of ECG signals [5], but not commonly as features for classification of arrhythmia because they mainly represent the

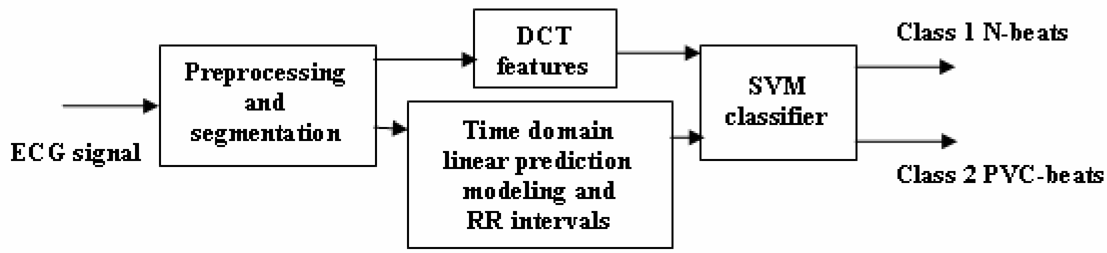

Figure 2: block diagram of the algorithm

shape (morph) of the signal and return a little information from time domain characteristics of the signal.

In this paper, we construct the feature vector with the aggregation of DCT coefficients and some time domain features like linear prediction modeling in each segment and R-R interval between segments.

By exact segmentation of the signal and tuning the SVM classifier parameters, we obtain up to 98.5% accuracy in classification that is comparable to one reported in [6] using advanced signal processing techniques such as higher order statistics. Figure 2 shows the block diagram of the implemented algorithms.

2. SEGMENTATION

The first stage in the ECG signal recognition system is segmenting the signal into pieces. To do this the recognition system must first locate fiducial points and then calculate the features from these segments. QRS is the main part of a heart beat and various methods have been used to detect QRS points in the ECG signal [7].

In this paper we use the method described in [8] because of its robustness to noise and accuracy. This method uses a linear and nonlinear scheme to provide input to the QRS detector decision section. The ECG signal is first filtered by a band pass filter to delete the signal noise and extract the desired components of the signal. The cleaned ECG is then followed by differentiation, squaring and time averaging window. With a proper threshold, we can detect the R point of the preprocessed signal. Detection of Q and R points is done by searching local minimum and local maximum in the left side of the R point by choosing a reasonable window search. Detection of the S and T points is done by the same method in the right side of the R point. We consider the interval between the P point and the T point as one beat segment.

3. FEATURE EXTRACTION

In [6] two variations of feature vector construction are examined. One is based on higher order statistics (HOS) of QRS segments and the other is based on hermit basis functions expansion (HBFE). In the first approach features are extracted from each QRS segment vector using 2^{nd} ,3^{nd} and 4^{th} order cumulants for five different time lags evenly distributed in the QRS segment and two classic RR interval variations.

In the second approach, the coefficients of the hermit basis function expansion and two classic RR interval variations are used as the feature vector. Equation 1 shows this expansion:

$$x(t) = \sum_{n=0}^{N-1} c_n \Phi_n(t,\sigma) \qquad (1)$$

Where c_n are the expansion coefficients and $\Phi_n(t,\sigma)$ is the hermit basis function of n^{th} order with parameter σ.

Computation of these coefficients requires SVD decomposition and pseudo-inverse technique, which is computationally expensive.
Both of these approaches try to include shape and time dynamics of each beat implicitly.

In this paper, we consider this information explicitly in the feature vector that leads to simpler and faster construction of the feature vector. We construct the feature vector such that it includes three types of information: first it must include the shape information of the beat, second it must include the intra beat time domain information for each beat and third it must have information about inter beats time domain variations.

Discrete Cosine Transform (DCT) is widely used for ECG signal compression. In this paper, we propose the coefficients of the DCT transform as the elements of feature vector. These features can approximate the shape of the signal. We use the first 10 elements of the DCT transform as the shape information for feature vector.

Physicians use not only the shape of the heart beat but also the duration and beats time intervals to diagnose the diseases, therefore it is necessary to include some time domain features to the feature vector.

Time domain modeling like linear predictive and ARMA modeling try to find the coefficients of a digital rational transfer function that approximate a given time domain impulse response. We chose linear prediction to model the intra beat variations.

Linear prediction modeling is a time domain based modeling that assumes each sample of the signal $x(k)$ is a linear combination of the past n signal samples, i.e. it can be linearly predicted from the past samples, and the coefficients are constant. Equation 2 shows this concept.

$$x(n) = a_1 x(n-1) + a_2 x(n-2) + \dots + a_k x(n-k) \quad (2)$$

We use four coefficients to consider intra beat time domain information in the feature vector.

To consider the time variations between inter beats, we use two features which are usually used: one element is the instantaneous RR interval of the beat which is computed as the distance between the QRS peaks of the present and previous beat, and the other element representing the average of RR intervals between the last 10 beats.

Therefore, a 16-element vector represents each beat.

4. SVM CLASSIFIER

The SVM is known as a good classifier because of its generalization performance [6]. Since the variations between the betas belonging to the same class in heart beats are wide, SVM is a good candidate for the classifier. Basically, the SVM is a linear machine working in the high dimensional feature space, formed by the nonlinear mapping of the n-dimensional input vector x into a k-dimensional, $(k > n)$ feature space through the use of a mapping $\phi(x)$.

A classification task using SVM usually involves training and testing the data. Given a training set of instance-label pairs $(x_i, y_i), i = 1,2,\dots,l$ where $x_i \in R^n$ is the feature vector and $y_i \in \{1,-1\}$ is its corresponding class label, the support vector machine requires the solution of the following optimization problem:

$$\min_{w,b,\xi} \; \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

$$\textit{subject to} \quad y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0$$

Here the training vectors x_i are mapped into a higher dimensional space by defining the function Φ . The SVM finds a linear separating hyper plane with the maximal margin in the higher dimensional space. $C \succ 0$ is the penalty parameter of the error, and $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is called the kernel function. The most commonly used kernel function is the radial basis function (RBF): $k(x_i, x_j) = \exp(\gamma \left\| x_i - x_j \right\|^2), \gamma > 0.$

We use the SVM classifier with the RBF kernel, and adjusting the γ parameter to get the best results in test data classification. Due to different input features, normalization of the feature vector elements is necessary. For this purpose each element of the feature vector is normalized according to the mean and variances of the corresponding elements in the same type beats. This normalization is also necessary for SVM classifier to work.

5. RESULTS

For the numerical experiments, we use the ECG data from the MIT BIH arrhythmia database [9]. We use different files of the database to consider the variation of the beats for different people. The total number of normal extracted beats is 3586 and the total number of PVC beats is 1232. In the training phase of the SVM classifier, we use 60% of the available data and test the classifier with the other 40%. The label +1 is used for normal beats and the label -1 is used for PVC beats. Results are tabled in tables 1.a and 1.b for both the training and testing data. We use confusion matrix to show the performance of the classifier. The element a_{ij} of this matrix represents the percentage of classifying the class i samples to class j. The first row of this matrix is for class 1, i.e. normal beats, and the second row is for class 2, i.e. PVC beats.

We compare our results to other approaches described in section 3. Table 2 shows the classification error for the proposed and the two other approaches. The proposed method has better classification error rate than the HOS method for the testing data. Although the classification error rate of hermit basis function expansion method is better than the proposed approach, however our algorithm is simpler and faster in computing the feature vector because the HBFE method requires SVD decomposition and pseudo-inverse computation.

Table 1.a: results of algorithm for training data

Classification rate = 99.7%		
class	1	2
1	99.8%	.2%
2	.3%	99.7%

Table 1.b: results of algorithm for testing data

Classification rate = 98.5%		
class	1	2
1	98.7%	1.3%
2	2.4%	97.6%

Table 2: classification error rate in percent for different methods

Method	training		testing	
	N	PVC	N	PVC
HOS	.85	2.04	2.85	2.75
HBFE	.5	1.02	1.5	1.7
proposed	.2	.3	1.3	2.4

6. CONCLUSION

This paper proposed a SVM based classifier for recognition of the normal and PVC heart beats. Three types of features for each beat are extracted and used as a feature vector. 10 DCT coefficients to consider the shape of beat, 4 linear prediction coefficients to consider the intra beat time domain information and 2 inter beat parameter extracted from RR intervals are used as a feature vector. These features are classified with the SVM classifier. Numerical experiments are done on MIT BIH Arrhythmias database and the classification accuracy is 98.5% for test data, which shows the accuracy of the implemented algorithm.

7. REFERENCES

[1]V.X. Afonso, W.J. Tompkins, "Detecting ventricular fibrillation", *IEEE , Engineering in Medicine and Biology Magazine,* Vol. : 14 , Issue: 2 , March-April 1995 Pages:152 - 159

[2] D. Strauss, G. Steidl, J. Jung, "Arrhythmia detection using signal-adapted wavelet preprocessing for support vector machines", *Computers in Cardiology 2001,* 23-26 Sept. 2001, Pages: 497 – 500

[3]M.Lagerholm, C.Peterson, G.Braccini, L.Ede nbrandt, L.Sornmo,"Clustering ECG Complexes Using Hermite Functions and Self-Organizing Maps",*IEEE Trans. On Biomed. Eng.,* vol.47,no.7,JULY2000

[4]Y. H. Hu, W. Tompkins, J. L. Urrusti, and V. X. Alfonso, "Applications of artificial neural networks for ECG signal detection and classification," *Electrocardiology*, vol. 24, pp. 123–129, 1994.

[5]V.A. Allen, J. Belina, "ECG data compression using the discrete cosine transform (DCT)", *Proc. Of Computers in Cardiology 1992. ,* 11-14 Oct. 1992, Pages: 687 – 690

[6] S. Osowski, L.T. Hoai, T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition", *IEEE Trans. on Biomed. Eng.,* Vol: 51 , Issue: 4 , April 2004, Pages:582 – 589

[7] B.-U.Kohler, C. Hennig, R.Orglmeister, "The principles of software QRS detection", *IEEE , Engineering in Medicine and Biology Magazine,* Vol: 21 , Issue: 1 , Jan.-Feb. 2002 Pages:42 - 57

[8] P. S. Hamilton, W.J. Tompkins, "Quantitative Investigation of QRS Detection Rules using the MIT/BIH Arrhythmia database," *IEEE Trans. On Biomed. Eng., vol. BME-33,* no9,pp. 1157-1165,1986

[9]http://www.physionet.org/physiobank/database/mitdb/