Automatic Classification of Arrhythmic Heartbeats using the Linear Prediction Model

Chun-Cheng Lin[1], Weichih Hu[2], Chun-Min Yang[1]

[1]Department of Electrical Engineering, National Chin-Yi University of Technology, Taichung, Taiwan
[2]Department of Biomedical Engineering, Chung Yuan Christian University, Chung Li, Taiwan

Abstract

This study developed an automatic heartbeat classification system based on the morphological features extracted using the first-order linear prediction model with two optimal filter coefficients and the RR interval features normalized by the heart rate of individual patient to reduce the effects of inconsistent heart rates among patients. Three heartbeat classes, normal beats, supraventricular ectopic beats and ventricular ectopic beats obtained from the MIT-BIH Arrhythmia Database, were used to test the performance of the proposed method. The ECG data were divided into training and testing datasets, each containing about 50,000 heartbeats. The training dataset was first used to establish the optimal linear discriminant classifier, and then the testing dataset was applied to evaluate the classification performance. The study results demonstrate that the sensitivity and positive predictive value of the proposed method were 88.7% and 99.4% for normal beats, 79.5% and 30.1% for supraventricular ectopic beats, and 88.6% and 57.7% for ventricular ectopic beats, respectively. If the RR interval features without normalization were used, the sensitivity and positive predictive value for supraventricular ectopic beats decreased to 62.5% and 24.0%, respectively.

1. Introduction

Automatic classification of heartbeats is an essential technique for identifying arrhythmic beats from large amounts of data in long-term ambulatory ECG recordings. It can also provide quantification information for the diagnosis of the risk of arrhythmias or sudden cardiac death such as the presence of ventricular premature beats and nonsustained ventricular tachycardia, and for further inspection, e.g. long-term heart rate variability and heart rate turbulence.

Several automatic heartbeat classification systems have been proposed by previous studies using various ECG features and classification methods for classifying arrhythmic beats. The ECG features include morphology [1-4], heart-beat intervals [1], RR intervals [1-4], higher order statistics [5], Hermite basis functions [5], and frequency-based features [6]. The classification methods include the linear discriminant classifier [1-3], quadratic classifier [2], patient-adapting heartbeat classifier [4], support vector machines [5], and back propagation neural networks [7]. However, one of the limitations of automatic heartbeat classification is the variation in the morphologies of ECG waveforms, not only among different patients or patient groups, but also within the same patient. Furthermore, the RR intervals involved in most of the previous studies are not only affected by the presence of arrhythmic beats, but also by the differences in heart rates among patients.

To further improve the classification performance of arrhythmic beats, this study developed an automatic heartbeat classification system. This system is based on the morphological features extracted using the linear prediction model [8] and the RR interval features normalized by the heart rate of individual patient to reduce the effects induced by the inconsistent heart rates among patients.

2. Methods of heartbeat classification

2.1. ECG data

All the ECG data used in this study were obtained from the MIT-BIH Arrhythmia Database [9] which contains common and life-threatening arrhythmic heartbeats. The MIT-BIH Arrhythmia Database contains 48 recordings of two-channel ambulatory ECG recordings with a length of 30 minutes, a sampling frequency of 360 HZ, and 11-bit resolution over a 10 mV range. In most recording, the upper lead is a modified limb lead II (MLII), while the lower lead is usually a modified lead V1 (occasionally V2 or V5, and in one instance V4). There are over 109,000 beats that are individually labelled as one of 15 possible heartbeat types. In accordance with the standards recommended by the Association for the Advancement of Medical

Instrumentation (AAMI) [10], the four recordings containing paced beats were removed from this study, and the remaining recordings were divided into training and testing datasets. The training dataset was used to establish the optimal linear discriminant classifier, and then the testing dataset was applied to evaluate its performance. The MIT-BIH heartbeats were classified into class N, consisting of the normal and bundle branch block beats, class S, consisting of supraventricular ectopic beats, and class V, consisting of ventricular ectopic beats, according to the AAMI recommendations [10]. Table 1 lists the heartbeat classes and numbers of the training and testing datasets.

2.2. Preprocessing

The preprocessing was to remove the high-frequency noise and baseline drift of the input ECG data before the feature extraction. The high-frequency noise was removed by a second-order low-pass filter with the z transfer function as follows:

$$H(z) = \frac{1}{36}\frac{(1-z^{-6})^2}{(1-z^{-1})^2} . \tag{1}$$

After removing noise, two median filters were designed for baseline estimation. The first median filter with a width of 200 ms was used to remove the QRS and P waves. The second median filter with a width of 600 ms was then applied to remove the T wave. After removing the QRS, P and T waves, the baseline drift can be estimated from the output of the second median filter, and then can be removed by subtracting the estimated baseline drift from the input ECG signal.

2.3. RR interval features

The RR interval was defined as the interval between successive R waves. There are four RR interval features involved in this study, defined as follows:

(a) Previous RR interval: the interval between a given R wave and the previous R wave.
(b) Post RR interval: the interval between a given R wave and the following R wave.
(c) Averaged 1-min RR interval: the mean of the RR intervals for one minute ECG recordings.
(d) Averaged 20-min RR interval: the mean of the RR intervals for twenty minute ECG recordings.

Table 1. Heartbeat classes and numbers of datasets.

Datasets	Heartbeat classes			Total
	N	S	V	
Training	45,824	943	3,787	50,554
Testing	44,218	1,836	3,219	49,273
Total	90,042	2,779	7,006	99,827

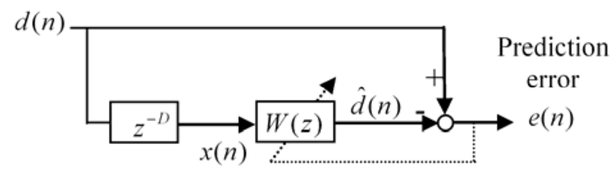

Fig. 1 Block diagram of a linear prediction model.

The four RR interval features were further normalized by the heart rate of individual patient to reduce the effects induced by the inconsistent heart rates among patients.

2.4. Morphology features extracted using the linear prediction model

Figure 1 is a block diagram of the linear prediction model for modeling the input QRS wave, where D is the prediction depth and $W(z)$ denotes the z-transform system function of a Wiener filter. The desired input $d(n)$ is the input QRS wave, and the input reference signal $x(n)$ is the delayed version of the input QRS wave, $x(n) = d(n-D)$. The prediction output of the Wiener filter with order $M-1$ can be represented as

$$\hat{d}(n) = x(n) \otimes w(n) = \sum_{i=0}^{M-1} w(i)x(n-i) , \tag{2}$$

where $\otimes$ denotes the operation of convolution sum and $w(i)$ for $i = 0, 1, \Lambda , M-1$ are the filter coefficients. The Wiener filter design problem requires finding the filter coefficients, $w(i)$, that minimize the mean-square error

$$\xi = E\{|e(n)|^2\} = E\{|d(n) - \hat{d}(n)|^2\} . \tag{3}$$

The optimal filter coefficients can be derived from the Wiener-Hopf equations [8] as follows:

$$\mathbf{R}_x \mathbf{w}_o = \mathbf{r}_{dx} , \tag{4}$$

where $\mathbf{R}_x$ is an $M \times M$ autocorrelation matrix of the reference input $x(n)$, $\mathbf{w}_o$ is an $M \times 1$ vector of the optimal filter coefficients, and $\mathbf{r}_{dx}$ is an $M \times 1$ vector of the cross-correlations between the desired input $d(n)$ and the reference input $x(n)$. This study introduces General Levinson Recursion [8] to recursively solve the Wiener-Hopf equations which are a set of Hermitian Toeplitz equations of the form given in Eq. (4). The optimal filter coefficients $\mathbf{w}_o$ of the linear prediction model with prediction depth D = 1 are applied as the morphological features of the QRS waves.

2.5. Linear discriminant classifier

A linear discriminant classifier [1] was optimized using the RR interval features and the optimal filter coefficients of the linear prediction model from the training dataset to classify the classes of N (normal and bundle branch block beats), S (supraventricular ectopic beats), and V (ventricular ectopic beats).

Assume total has g classes, and $\mathbf{X}$ denotes the feature vector. The classification rule for a given feature vector is based on the calculation of the posterior probability defined as follows:

$$P(i\,|\,\mathbf{X}) = \frac{e^{d_i(\mathbf{X})}}{\sum_{j=1}^{g} e^{d_j(\mathbf{X})}} , \tag{5}$$

where $d_i(\mathbf{X})$ denotes the discriminant value and is defined as follows:

$$d_i(\mathbf{X}) = \frac{1}{2}(\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i) + (\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{X}) + \log(\pi_i) , \tag{6}$$

where $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}$, and π_i denote the mean vector, covariance matrix, and prior probability, respectively. The mean vector and covariance matrix are defined as follows:

$$\boldsymbol{\mu}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} \mathbf{X}_{ni} , \tag{7}$$

and

$$\boldsymbol{\Sigma} = \frac{1}{N_i - g} \sum_{i=1}^{g} \sum_{n=1}^{N} (\mathbf{X}_{ni} - \boldsymbol{\mu}_i)(\mathbf{X}_{ni} - \boldsymbol{\mu}_i)^T . \tag{8}$$

where $\mathbf{X}_{ni}$ denotes the nth feature vector in the ith class, and T represents the transport operation. A given feature vector is classified into a class that has the maximal posterior probability among all classes.

2.6. Performance parameters

Two local performance parameters are defined to evaluate the performance of the automatic heartbeat classification system for class i including the sensitivity, Se_i , and positive predictive value, P_i^+ , defined as $Se_i = \frac{n_{ii}^T}{N_i}$, and $P_i^+ = \frac{n_{ii}^T}{P_i}$, where n_{ii}^T is the number of correctly classified heartbeats, N_i is the total number of heartbeats for class i, and P_i is the number of heartbeats classified as class i.

Three global performance parameters were defined to evaluate the classification performance for all classes including global sensitivity, Se_i , global positive predictive value, P^+ , and global accuracy, Acc, respectively defined as $Acc = \frac{1}{N_T} \sum_{i=1}^{g} n_{ii}^T$, $Se = \frac{1}{g} \sum_{i=1}^{g} Se_i$, and $P^+ = \frac{1}{g} \sum_{i=1}^{g} P_i^+$, where N_T is the total number of heartbeats in the dataset.

3. Results

Figures 2 and 3 illustrate the prediction outputs (dashed line) for a normal QRS wave (solid line) and a ventricular ectopic QRS wave (solid line) using a first-order linear prediction model with two optimal filter coefficients, $w(0)$ and $w(1)$, respectively. It is shown that the prediction output can estimate most of the morphology of the input QRS wave, and the morphological features extracted by the two optimal filter coefficients can distinguish the normal QRS wave from the ventricular ectopic QRS wave (1.92 vs. 1.76 for $w(0)$, and -1.96 vs. -0.78 for $w(1)$).

Table 2 shows the classification results for the testing dataset using RR interval features and two optimal coefficients of a first-order linear prediction model. All of the classification parameters can be improved by using the normalized RR interval features in comparison with the use of the RR intervals without normalization. Especially, the local sensitivity and positive predictive value for class S can be increased significantly from 62.5% to 79.5%, and from 24.0% to 30.1%. The global parameters can be improved from 86.5% to 88.4% for accuracy, from 79.5% to 85.6% for sensitivity, and from 58.9% to 62.4% for positive prediction value.

4. Discussion and conclusions

This study demonstrates an automatic heartbeat classification system based on the normalized RR interval features and morphological features extracted by a linear prediction model. The RR interval features are mainly applied to distinguish supraventricular ectopic beats from normal beats because the shapes of the QRS waveforms are very similar. To reduce the effects caused by the inconsistent heart rates among patients, the RR interval features are normalized by the heart rate of individual patient in this study. The study results also show that the use of the normalized RR intervals results in a substantial improvement in the local sensitivity and positive prediction value of the classification of the supraventricular ectopic beats. The morphological features are mainly used to identify the ventricular ectopic heartbeats because their waveform shapes are different with those of the normal and supraventricular ectopic heartbeats. The combinations of the proposed morphological features and the RR intervals without and with normalization can reach accuracy of 86.5% and 88.4%, respectively.

In conclusion, the proposed automatic heartbeat classification system based on the linear prediction modeling is demonstrated to be useful in terms of identifying normal, supraventricular ectopic, and ventricular ectopic heartbeats. Future works would be focused on the improvement of the low positive prediction value for the identification of the supraventricular and ventricular ectopic heartbeats.

Acknowledgements

This research was financially supported by the National Science Council under Grant No. NSC 101-2320-B-167-001-.

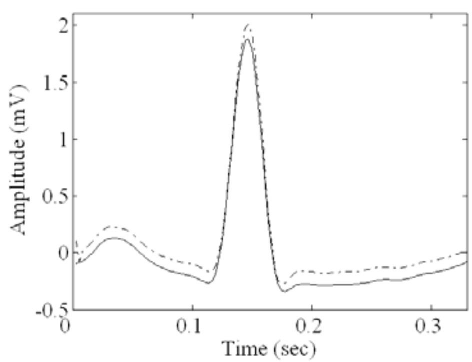

Figure 2. A normal QRS wave (solid line) and the prediction output (dashed line) ($w(0) = 1.92$, $w(1) = -1.96$)

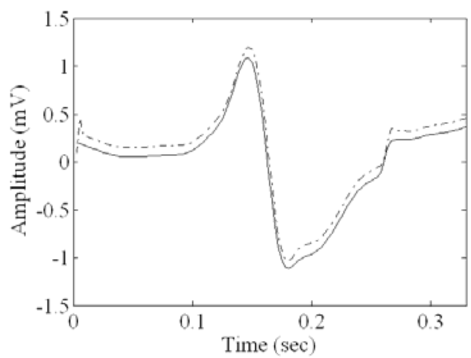

Figure 3. A ventricular ectopic QRS wave (solid line) and the prediction output (dashed line) ($w(0) = 1.76$, $w(1) = -0.78$)

Table 2. Classification results for the testing dataset.

ECG Features	Local Parameters (%)						Global Parameters (%)		
	N		S		V				
	Se	P^+	*Se*	P^+	*Se*	P^+	*Acc*	*Se*	P^+
RR + LP model	87.4	98.5	62.5	24.0	88.4	54.4	86.5	79.5	58.9
Normalized RR + LP model	88.7	99.4	79.5	30.1	88.6	57.7	88.4	85.6	62.4

References

[1] Chazal PD, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans Biomed Eng 2004;51:1196-206.

[2] Llamedo M, Martinez J. An ECG classification model based on multilead wavelet transform features. Comput Cardiol 2007;34:105-8.

[3] Llamedo M, Martinez JP. Heartbeat classification using feature selection driven by database generalization criteria. IEEE Trans Biomed Eng 2011;58:616-25.

[4] Chazal PD, Reilly RB. A patient-adapting heartbeat classifier using ECG morphology and heartbeat interval features. IEEE Trans Biomed Eng 2006;53:2535-43.

[5] Park K, Cho B, Lee D, Song S, Lee J, Chee Y, Kim I, Kim S. Hierarchical support vector machine based heartbeat classification using higher order statistics and hermite basis function. Comput Cardiol 2008;35:229-32.

[6] Senhadji L, Carrault G, Bellanger JJ, Passariello G. Comparing wavelet transforms for recognizing cardiac patterns. IEEE Eng Med Biol Mag 1995;14:167-73.

[7] Osowski S, Linh TL. ECG beat recognition using fuzzy hybrid neural network. IEEE Trans Biomed Eng 2001;48:1265-71.

[8] Hayes MH. Statistical Digital Signal Processing and Modeling. John Wiley & Sons Inc 1996.

[9] Mark R, Moody G. MIT-BIH Arrhythmia Database [Online] 1997. http://ecg.mit.edu/dbinfo.html

[10] Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms, 1998. Association for the Advancement of Medical Instrumentation.

Address for correspondence.

Chun-Cheng Lin
Department of Electrical Engineering, National Chin-Yi University of Technology
No.57, Sec 2, Jhongshan Rd., Taiping Dist., Taichung 41170, Taiwan
cclin@ncut.edu.tw