PATTERN RECOGNITION OF CARDIAC ARRHYTHMIAS BASED ON MULTIVARIATE AUTOREGRESSIVE MODELING

Dingfei Ge*, Zhegen Zhang
Information & Electrical Engineering Department,
Zhejiang University of Science and Technology, Hangzhou 310012, China

Abstract—Computer-assisted automatic diagnosis will play an important role in diagnosis and treatment of critical ill patients. Multivariate autoregressive modeling (MAR) has been performed on two-lead ECG signals. MAR coefficients and K-L transformation of MAR coefficients have been used as ECG features for classification. Five types of ECG signals were obtained from MIT-BIH database, namely normal sinus rhythm, atria premature contraction, premature ventricular contraction, ventricular tachycardia, and ventricular fibrillation. A quadratic discriminant function (QDF) based classification algorithm was employed in this study. The results show MAR coefficients produced slightly better results than K-L transformation of MAR coefficients. The accuracy of classification based on MAR coefficients was 96.6% to 99.3%.

Key words: ECG signals, Multivariate Autoregressive Modeling, Quadratic discriminant function, Classification

INTRODUCTION

Automatic discrimination of rhythms in Electrocardiogram (ECG) signals will play an important role in diagnosis and treatment of critical ill patients. Those arrhythmias like ventricular tachycardia (VT) and ventricular fibrillation (VF) are most dangerous. Other arrhythmias including premature ventricular contraction (PVC) and atria premature contraction (APC) are not so lethal, but are also important for diagnosing the heart diseases. In fact, even small improvements to the reliability of such algorithms will translate into lives being saved. Various studies have been done for classification of various cardiac arrhythmias, such as correction waveform analysis [1], total least squares based Prony modeling algorithm [2], and complexity measures etc.[3]. Generally, these techniques classify two or three arrhythmias only and are difficult to implement and compute. Multivariate autoregressive modeling (MAR) has been widely used to model bio-signals for analysis. Two-lead ECG signals contain more information than one-lead ECG signals. Thus, the accuracy and reliability of detecting various arrhythmias can be improved by using two-lead ECG signals.

The purpose of the work described in this paper is to model two-lead ECG signals to improve cardiac arrhythmias classification accuracy using MAR modeling. MAR modeling was performed on ECG data from normal sinus rhythm (NSR), APC, PVC, VT and VF. The MAR coefficients were utilized as ECG features for classification. Quadratic discriminant function (QDF) based classification was performed in various stages in current study.

1 METHODS

1.1 ECG Data Acquisition, Filtering and Segmentation

The frequency of the selected data of NSR, PVC and APC in this study was 360HZ, and the frequency of the selected VT and VF was 250Hz. The data including VT and VF was subsampled in order that all the ECG signals in the analysis had a frequency of 360 Hz for ECG data filtering. After the filtering, subsampling processes were applied to all the two-lead ECG data again in order to obtain a sampling frequency of 250Hz for MAR modeling. An integer band-pass filter was used to filter the ECG signals. The lower frequency passband and upper frequency passband are 2Hz and 20 Hz respectively. The R peaks of ECG were detected using Tompkin's algorithm. In current study, the sample size of the various segments was 0.9 seconds. 0.3 seconds before R peak and 0.6 seconds after R peak were picked for MAR modeling. Total 250 segments each from the five classes were selected for analysis in this research.

1.2 MAR Modeling

A two-channel MAR model of order M is give by [4]

$$HR1(n) = \sum_{k=1}^{M} HR1(n-k)a_{11}(k) + \sum_{k=1}^{M} HR2(n-k)a_{12}(k) + e_1(n) \quad (1)$$

$$HR2(n) = \sum_{k=1}^{M} HR2(n-k)a_{22}(k) + \sum_{k=1}^{M} HR1(n-k)a_{21}(k) + e_2(n) \quad (2)$$

where $HR1(n)$ and $HR2(n)$ represent ECG lead time series respectively, $e_1(n)$ and $e_2(n)$ represent unknown, zero mean and uncorrected random variables respectively, a_{ij}'s are the MAR model coefficients.

Burg's algorithm was used to compute the MAR coefficients. The criterion used to evaluate the model order selection was the sum-squared error (*SSE*) in this work. The *SSE* for MAR modeling is given by

$$SSE = \sum_{i=1}^{2} \sum_{j=M+1}^{N-M} e_i(j) \quad (3)$$

where N is the number of the sample points of each segment used for modeling (0.9seconds).

1.3 ECG Feature Extraction for Classification

MAR coefficients, the Karhunen-Loeve (K-L) transformation of the MAR coefficients (K-L MAR coefficients) were used to represent the ECG segments. The number of the MAR parameters representing an ECG segment was $4M$. Karhunen-Loeve (K-L) transformation of the MAR feature vectors was computed to reduce the number of features. The K-L transformation in this study was performed as follows [5]:

(1) Calculate the covariance matrixes of each class Σ_i. (2) Calculate the within-class scatter matrix S of these classes.

$$S = \sum_{i=1}^{C} P_i \Sigma_i \quad (4)$$

where P_i is the prior probability of ω_i, $\omega_i = 1/C$ in current study, C is the total number of the classes in this research. (3) Calculate the eigenvalue r_i's (i=1,2,…) and eigenvectors of the within-class scatter matrix S. (4) The set of m eigenvectors which correspond to the m largest eigenvalues was chosen to transfer the original data, the choice of the number of eigenvectors in this study was given by the index i for which $r_i/r_{max} \leqq 0.001$, r_i's are in decreasing order. (5) Generate the K-L transformation by projecting each $4M$-dimentional pattern onto these chosen eigenvectors.

1.4 QDF-Based Classification

In the current study, the cardiac arrhythmias were classified by using a stage-by-stage QDF-Based algorithm. The QDF is given by [5]

$$y_i = \beta_0 + \sum_{k=1}^{d} \beta_k x_k + 2\sum_{m=1}^{d-1} \sum_{n=m+1}^{d} \beta_{mn} x_m x_n + \varepsilon_i \quad (5)$$

and the matrix form is given by

$$y_i = X_i \beta + \varepsilon_i \quad (6)$$

where $[x_1, x_2, \ldots, x_d]$ represents an ECG feature vector in this study; y_i is an observe response, e_i is the QDF error; X_i is a $(d(d+3)/2+1)$-dimensional row vector, β is a $(d(d+3)/2+1)$-dimensional column vector, that is:

$$X_i = [1, x_1, x_2, \ldots, x_d, x_1^2, x_2^2, \ldots, x_d^2, 2x_1x_2, 2x_1x_3, \ldots, 2x_1x_d, 2x_2x_3, 2x_2x_4, \ldots, 2x_2x_d, \ldots, 2x_{d-1}x_d]$$

$$\beta = [1, \beta_1, \beta_2, \ldots, \beta_d, \beta_{11}, \beta_{22}, \ldots, \beta_{dd}, \beta_{12}, \beta_{13}, \ldots \beta_{1d}, \beta_{23}, \beta_{24}, \ldots, \beta_{2d}, \ldots, \beta_{(d-1)d}]^T$$

The ECG feature vector of a particular ECG segment was mapped to a response ("1" or "–1"). Assume the total number of the ECG segments used for classification at a particular stage is D. The following equation can be given

$$\widetilde{Y} = A\beta + E \quad (7)$$

where $\widetilde{Y} = [y_1, y_2, \ldots, y_D]^T$ is a D-dimensional column vector of the observed responses, and made up of "1" and "-1", which correspond to different classes respectively, $A=[X_1, X_2, \ldots, X_D]^T$ is a $D \times (d(d+3)1/2+1)$ matrix, $E=[e_1, e_2, \ldots, e_D]^T$ is a D-dimensional column vector of the errors.

The least squares estimator is given by

$$\beta = (A^T A)^{-1} A^T \widetilde{Y} \quad (8)$$

The discriminant function of the estimator is given by

$$YI = X_i \beta \quad (9)$$

The various stages of the classification for MAR and K-L MAR coefficients are shown in figure 1. Euclidean distance between the mean feature vectors of different classes was computed to determine the groupings of the classes at each stage in order to perform the stage -by -stage classification. During the training phase, the estimator β was computed as the equation of (8) using the selected training sets at each stage of classification. During the testing phase, the output response at each stage of the classification (*Y1* in stage1, *Y2* in stage2, etc.) was computed as the equation of (9) using the feature vectors and the estimated β. A threshold value of zero was used to classify the output response as belonging to a group at a particular stage.

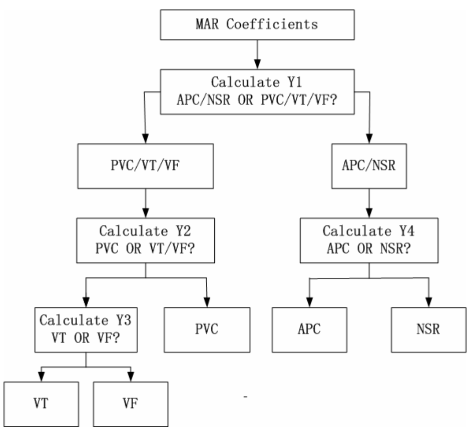

Figure1. QDF-based classification algorithm

2 Results

2.1 MAR Modeling Results

The modeling results show the *SSE* for various arrhythmias decreases initially with the model order *M*, but remains almost constant for model order greater than or equal three. However, A MAR model of order four was selected for extracting the features in this research. This is because more detail can be incorporated into the model order that might be missing from a lower-order model.

A fitter scalar AR model of order 4 also was found to model single-lead ECGs using *SSE*. This is consistent with the other researches on the AR model order selection [6].

2.2 Classification Results

The Euclidean distances between the different class centers was computed to determine the grouping. It was found that VT/VF/PVC, APC and NSR form one group respectively due to small center distance within the same group and large center distance between different groups, and VT/VF form a subgroup within group VT/VF/PVC according to the Euclidean distance. Therefore, VT/VF/PVC was separated from APC and NSR in stage one (*Y1*). The membership of VT/VF /PVC was defined as " +1" , and the membership of APC and PVC was defined as "-1" . The least squares estimator β was computed as the equation of (8). The output response *Y1* was computed as the equation (9). The value *Y1* was used to determine the classes. Similarly, in the second stage (Y2), PVC was differentiated from VT/VF. Stage three (*Y3)* and four (*Y4*) were used to distinguish between NSR, APC, VT and VF as shown in Figure 1.

One hundred and fifty cases each from the five classes were random selected to estimate β in training phase, and the remaining was used for testing. The average classification results based on MAR coefficients on test data are given in Table 1.

The number of the eigenvectors was chosen to be 8 according to the choice criterion of eigenvectors. The 8-dimensional feature vectors based on K-L transformation were trained and tested as the same way as the classification based on the MAR coefficients, the classification accuracy based on K-L MAR coefficients was given in the Table 2.

For comparison between two-lead and one-lead ECG signals modeling for classification, the classification accuracy based on scalar AR coefficients representing single-lead ECG segments was given in table3.

Table1 Classification accuracy based on MAR coefficients

Classes	NSR	APC	PVC	VF	VT
Accuracy	98.5%	96.6%	97.3%	99.3%	98.8%

Table2 Classification accuracy based on K-L MAR coefficients

Classes	NSR	APC	PVC	VF	VT
Accuracy	96.6%	96.5%	96.5%	96.3%	97.1%

3 Discussions

The objective of this study was to model two-lead ECG signals in order to improve the classification results using MAR modeling. The experiment results show that MAR model order of 4 was sufficient to model ECG signals for the purpose of classification.

Table3 Classification accuracy based on scalar AR coefficients representing single-lead ECG segments

Classes	NSR	APC	PVC	VF	VT
Accuracy	96.6%	94.7%	91.8%	94.5%	94.3%

The classification results show that the classification accuracy based on MAR coefficients was slightly higher than the one based on K-L MAR coefficients. Extra calculation was involved in K-L transformation of MAR coefficients. Therefore, this representation may not be worth considering for a real-time system. The single-lead ECG based classification accuracy is much lower than the two-lead ECG based one. Thus, MAR modeling based classification can improve the classification accuracy significantly.

Some of the proposed techniques use only a smaller number of arrhythmias than current study, for example, classify PVC and NSR using a fuzzy ARTMAP classifier, accuracy of 97% for PVC was achieved [7]. VT and VF were classified using a modified sequential probability ratio test algorithm, the accuracy of detecting VT and VF were 93% and 96% respectively [8].

The sample size of the segments was 0.9 seconds only, and it was 3 to 7 seconds for the complexity measure-based technique [3] and 5 to 9 seconds for the Prony modeling technique [2].

MAR model might not be suited to ECG signals under all conditions since MAR model is a linear model, Further work can try to use nonlinear parameter models that might better capture the non-stationary nature of the ECG. Other further work can try to increase to the number of ECG leads. Further work is in progress to try to include more other types of classes.

4 Conclusions

The classification results could be improved by using two-lead ECG signals. A MAR modeling and QDF based classification are suitable for real-time implementation for diagnosis purpose.

Reference

[1] S. A. Caswell, K. S. Kluge, and C. M. J. Chiang, " Pattern recognition of cardiac arrhythmias using two intracardiac channels," Proc. Comp. Cardiol., pp. 181-184, 1993

[2] S. W. Chen, " Two-stage discrimination of cardiac arrhythmias using a total least squares-based prony modeling algorithm," IEEE Trans. Biome. Eng., 47, pp. 1317-1326, 2000.

[3] X. S. Zhang, Y. S. Zhu, N. V. Thakor, and Z. Z. Wang, " Detecting ventricular tachycardia and fibrillation by complexity measure," IEEE. Trans. Biomed. Eng., 46, pp. 548-555, 1999.

[4] I. Korhonen. "Multivariate closed-model for analysis of cardiovascular dynamics," Methods of Information in Medicine, 36, pp. 264-7, 1997.

[5] K. Fukunaga, Introduction to statistical pattern recognition. Academic press limited, United States of America, 1990.

[6] D. F. Ge, N. Srinivasan, and S. M. Krishnan, "Cardiac arrhythmia classification using autoregressive modeling," Biomedical Engineering Online, 1:5, 2002.

[7] F. M. Ham and S. Han. " Classification of cardiac arrhythmias using fuzzy ARTMAP," IEEE. Trans. Biomed. Eng., 43, pp. 425-430, 1996.

[8] S.W. Chen, P. M. Clarkson and Q. Fan, " A robust sequential detection algorithm for cardiac arrhythmia classification," IEEE Trans. Biomed. Eng,43, pp.1120-1121, 1996.