A Statistical Feature Based Approach to Predicting Termination of Atrial Fibrillation

FM Roberts, RJ Povinelli

Marquette University, Milwaukee, WI, USA

Abstract

This paper presents a nonlinear signal classification approach to differentiating different atrial fibrillation termination stages; non-terminating (N), terminating in one minute (S), and terminating immediately (T) following the end of the recording. The nonlinear approach is based on Gaussian mixture models of reconstructed phase spaces of the last 2s of data in each recording. The removal of the ventricular component of the signal was removed by one of two methods: QRST averaging and subtraction and 4-9 Hz bandpass filtering of the recording.

The accuracy of the approach is 63.3 and 66.7% for differentiating N vs. T and 60 and 70% accuracy for differentiating S vs. T, for QRST subtracted and filtered data respectively. When the training data was augmented with 150 more training cases, the results improved to 66.7 and 80% for N vs. T and 70 and 70% for S vs. T.

An artifact was noted in the recordings that allowed a different set of criteria (the slope and percent of data remaining after last Q point at the end of each of the recordings) to accurately classify N vs. T at 80% and S vs. T at 85%.

1. Introduction

The Computers in Cardiology Challenge 2004 was to determine the stage of spontaneous atrial fibrillation (AF) from one minute ECG recordings. The challenge was set up to improve the understanding of the mechanism of spontaneous AF termination with the desire that this understanding may lead to improvements in the treatment of paroxysmal and sustained AF.

Atrial fibrillation is the most common arrhythmia, in the US alone approximately 2.2 million people have AF; with about 15% of all strokes occurring in people with AF [1]. AF is the single most responsible arrhythmia for hospitalization. Almost 34% of all hospitalizations attributed to arrhythmias are due to AF. AF also occurs in 20-40% of patients undergoing coronary artery bypass surgery. This postoperative AF increases the duration and cost of the hospitalization [2]. Thus, if this challenge leads to a better understanding of the mechanisms of AF termination and an improved treatment for AF, the cost associated with AF hospitalization may be reduced.

Our research has a theoretical basis from the work of Takens [3] and Sauer et al. [4]. This work shows that a time series of observations, sampled from a single state variable can be used to reconstruct a space that is topologically equivalent to the original system. The researchers have had success using reconstructed phase space models to differentiate multiple cardiac arrhythmias [5-7].

The construction of a RPS is the embedding of time lagged versions of the original time series. Given a time series $x=x_n$, $n=1 \ldots N$, an RPS matrix $\boldsymbol{X}$ of dimension d and time lag τ is defined by its row vectors:

$$\boldsymbol{x}_n = [x_{n-(d-1)\tau} \; . \; . \; . \; x_{n-\tau} \; x_n], \qquad (1)$$

where $n = (1+(d-1)\tau) \ldots N$. A row vector $\boldsymbol{x}_n$ is a point in the RPS.

The sufficient condition for topological equivalence is that d is greater than twice the box counting dimension of the original system [4]. When d is not known, as is the case for most real systems (i.e. ECG data), it may be estimated using the false nearest-neighbour technique [8], which calculates the percentage of points which are near because of projection rather than dynamics. In Takens' original work, $\tau=1$. However, in practice it has been found that the appropriate selection of the time lag can reduce the required RPS dimension. A common heuristic for determining time lag is to use the first minimum of the automutual information function [8].

The proposed classification algorithm is theoretically capable of differentiating between signals generated by topologically different systems. Our approach builds Gaussian Mixture Models (GMMs) of signal trajectory densities between signals using a Bayesian classifier. The challenge was approached as a blind box data mining task with only the cardiac knowledge that atrial activation is in the 4-9Hz range to differentiate the different AF termination classes.

2. Methods

As discussed above, our approach to classify the different AF termination stages is to build GMMs of the signal trajectory densities in an RPS and differentiate between the classes using a Bayesian classifier. This is done in three steps. The first step, data preprocessing, includes the normalization of the data, estimating the time lag and dimension of the RPS, and removing QRS complex information in the signal. The second step is learning the GMMs for each signal class. The final step is signal classification, which is done with a maximum likelihood Bayes classifier.

In this section the method used to classify the different stages of atrial termination will be discussed after a short discussion about the data.

2.1 Challenge Data

Three sets of data sampled at 128 Hz were provided for the challenge training data, with two different sets of data for the test sets. The recordings were all one minute long excerpted from 24 hour ECG recordings.

The training data had 30 recordings that were divided into three different classes (10 each) corresponding to when AF terminated with respect to the end of the recordings:

- Class N – Non terminating AF or at least 60 minutes prior to the termination of AF
- Class S – AF terminating one minute after the end of the recording
- Class T – AF immediately terminating after the end of the recording

There were two sets of test data which were patient independent of the training test sets. The first test set consisted of 30 recordings that were either Class N or Class T. The second test set consisted of 20 recordings of either Class S or Class T.

2.2 Supplemental Training Data

One hundred fifty one minute recordings were selected by the researchers to augment the original 30 records of training data. These supplemental recordings consisted of 50 recordings of each AF termination class. The records were excerpted from 24-48 ECG Holter recordings, all of which were taken from lead I. The supplemental training data was sampled at 128 Hz.

The supplemental data were selected to allow for a more rigorous training set. It was also needed to provide sampling of the heart's state variables from the same lead.

2.3 Data Preprocessing

Each data recording was normalized to zero mean and unit standard deviation. As RPS models are morphology based, the data needs to fit within the same amplitudes. Time lags were calculated for each normalized signal using the first minimum of the automutual information function [8]. The overall time lag of 11 was selected using the mode of the histogram of all time lags. The RPS dimension of 3 was calculated using the false nearest neighbor technique [8].

To remove the QRS influence in the signals, the QRS were removed by band filtering the signal at 4-9 Hz and by averaging the previous 10 QRST complexes and subtracting from the signal.

2.4 Learn Models

The second step of the approach was to learn a GMM probability distribution for each AF termination class. A different set of models were learned for the filtered signals from the QRST subtracted signals. This was done by creating an RPS using the time lag and dimension determined in the previous step and inserting all the signals for a particular class into this phase space as described in (1) above.

The parameters describing the mean and variance of the 16 mixtures in each GMM model are determined the Expectation-Maximization (EM) algorithm. A model was generated for each termination class.

2.5 Classification

The last step of the algorithm is to classify the test signals. Each test signal is embedded in a RPS and then the conditional likelihood of each class model is calculated. Using a Bayesian classifier the maximum likelihood AF termination class is determined. When determining the classification of the training data a leave-one-out approach was used in order not to have the data being tested in the GMM model. This was only done with the challenge supplied training data. It was not needed when using the supplemental data because it was different recordings.

3. Results

When using the challenge training data, the maximum accuracy for 17 out of 20 and that was for both N vs. T and S vs. T both were bandpass filtered data. The maximum accuracy for the test sets were 20 out 30 for N vs. T and 14 out 20 for S vs. T again this was for the filtered data.

When the challenge training data was augmented with 150 more recordings these accuracies increased to 18 out of 20 for N vs. T for the filtered data and 16 out of 20 for S vs. T for both filtered and QRST subtracted data for the training set. The test set accuracy also increase for the N vs. T to 24 out of 30 for the filtered data. The accuracy of the test case for S vs. T did not increase, however the accuracy of QRST subtracted data increased to 14 out of 20. The individual accuracies can be seen in Table 1.

Table 1. Results of training and test sets for modelling using 4-9 Hz bandpass filtered and QRST subtracted data. Using both challenge supplied training data and supplemental data.

	Training Data		*Testing Data*	
	N vs. T (20)	S vs. T (20)	N vs. T (30)	S vs. T (20)

Challenge Training Data	Filtered	17	17	20	14
	QRS subtraction	16	15	19	12
Supplemental Training Data	Filtered	18	16	24	14
	QRS subtraction	17	16	20	14

4. Discussion and conclusions

When the QRST is subtracted from the signal recording some information pertaining to the atrial activation must also be being removed. This causes a lower accuracy for the QRST subtracted data to that of the filtered data.

Since both QRST subtracted and filtered data used RPS models to classify the AF termination stage and RPS require that the data come from the same state variable, the accuracy of neither method may greatly improve if modifications to the method are made if we continue to use these provided test recordings. Because the recordings are not all taken from the same lead, and they may not have even been using lead I, which is the lead used in the supplemental data. With that in mind then, the 80 and 70% accuracy for N vs. T and S vs. T respectively are reasonable results. The same challenge should be applied using the supplemental data that we know is from lead I to see if using the RPS models to determine AF is a good model.

When the data was being observed, we noticed an artificial feature that could be used to classify the different AF termination classes. In several of the recordings, it was noted that the recordings were ending on or near the QRS complex; this was seen extensively in the classes N and S. An example of this can be seen in Figure 1. Notice that the first two plots are N and T respectively and the recording ended on or just after the QRS complex. In the third plot, there is more recording after the QRS complex, it seems to be the rest of the last heart beat.

With those two differences in mind two features were calculated. The slope of the last 5 points of all the challenge datasets were computed, and the percentage of points remaining after the last QRS fiduciary point to the average beat length for the previous 5 beats. When looking at the remaining percentage, Class T was greater than 0.55 for all of the training set. The slopes for the training set Class T ranged between -0.05 and 0.17. The slopes for Classes N and S ranged from -1.00 to 0.07 and -0.09 to 1.22 respectively.

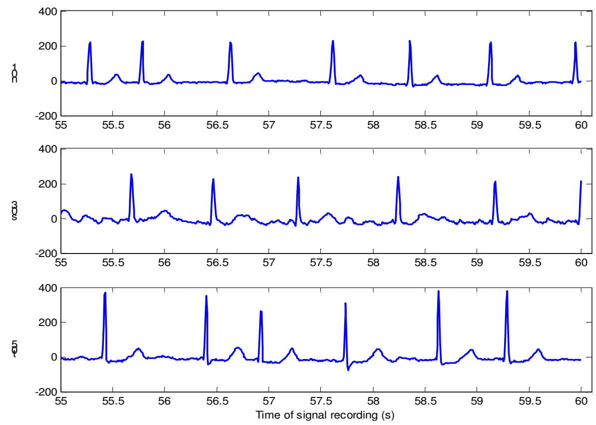

Figure 1. This figure shows the difference in the end of the recordings (n01, s03, and t05) between the three different AF termination classes. Note that both Class N and Class S may end the recording on or around a QRS complex.

These feature values from the immediately terminating AF class were used to differentiate the different signal recordings. The features were used separately and together to get a maximum training accuracy of 90 and 95% for N vs. T and S vs. T respectively. The accuracies of the test cases were 80 and 85% for N vs. T and S vs. T respectively. These accuracies were for a combination of the two features. The individual accuracies for both features can be seen in the Table 2.

Care should be taken when selecting datasets for such a challenge to not include artificial features that make accurate classifiers. A possible way to have avoided these artefacts would have been to verify that the recordings for the N and S classes ended just prior to the next QRS complex. Thus, the Class N would be labelled ending close to 60s after the end of the recording. Neither of these artificial features model the mechanisms of the termination of AF.

Table 2. Results of training and test set accuracies when using artificial features found in the supplied datasets; slope of last 5 points in recording, percent of beat remaining after last Q point, and a combination of the two.

		Training Data		*Testing Data*	
		N vs. T (20)	S vs. T (20)	N vs. T (30)	S vs. T (20)
Artifact	Slope	16	19	20	12
	%Beat	18	14	23	15
	Slope & Beat	18	19	24	17

Acknowledgements

An American Heart Association Pre-doctoral

Fellowship provided funding for this project. The authors would also like to thank Cardiac Evaluation Center (Milwaukee, WI) for supplying supplemental data for this research.

References

[1] Heart and Stroke Facts 2003. Dallas, TX: American Heart Association; 2003.
[2] Lip GYH, Mangrum JM, DiMarco JP. Atrial Fibrillation. In: Crawford MH, DiMarco JP, Paulus WJ, editors. Cardiology. Philadelphia: Mosby; 2004: p. 699-716.
[3] Takens F. Detecting strange attractors in Turbulence. Presented at Dynamical Systems and Turbulence, Warwick, 1980.
[4] Sauer T, Yorke JA, Casdagli M. Embedology. Journal of Statistical Physics 1991;65(3/4):579-616.
[5] Roberts FM, Povinelli RJ, Ropella KM. Identification of ECG Arrhythmias using Phase Space Reconstruction. In: DeRaedt L, Siebes A, editors. Principle and Practice of Knowledge Discovery in Databases (PKDD'01); 2001; Freiburg, Germany: Springer Verlag; 2001. p. 411-423.
[6] Povinelli RJ, Roberts FM, Ropella KM, Johnson MT. Are Nonlinear Ventricular Arrhythmia Characteristics Lost: As Signal Duration Decreases?. In: Computers in Cardiology; 2002; Memphis, Tennessee: IEEE; 2002. p. 221-224.
[7] Roberts FM, Povinelli RJ, Ropella KM. Rhythm Classification Using Reconstructed Phase Space of Signal Frequency Subbands. In: Computers in Cardiology; 2003; Thessaloniki, Greece: IEEE; 2003. p. 61-64.
[8] Kantz H, Schreiber T. Nonlinear Time Series Analysis. Cambridge: Cambridge University Press, 1997.

Address for correspondence

Felice M. Roberts
EECE Department, Marquette University
PO Box 1881, Milwaukee, WI 53201-1881, USA
Felice.Roberts@Marquette.edu

Richard J. Povinelli
EECE Department, Marquette University
PO Box 1881, Milwaukee, WI 53201-1881, USA
Richard.Povinelli@Marquette.edu