Statistical Analysis of RR Interval Irregularities for Detection of Atrial Fibrillation

A Ghodrati, S Marinello

Draeger Medical Systems, Inc, Andover, MA, USA

Abstract

We compare two Atrial Fibrillation (AFIB) detection methods from surface ECG, based on RR interval variability in a statistical framework. We obtain the histogram of normalized RR differences for AFIB and non-AFIB episodes using MIT-BIH Arrhythmia database. Two probability density functions (pdf) are employed to model the histograms: Gaussian and Laplace. We then use Neyman-Pearson (NP) detection approach to obtain criteria for AFIB detection. The performance of the two methods is compared using Receiver Operating Curves (ROC) over the different databases. The result shows that the Laplace pdf approximates the histogram of normalized RR differences better than the Gaussian pdf and leads to better AFIB detection performance.

1. Introduction

AFIB is the most common arrhythmia in clinical practice accounting for approximately one-third of hospitalizations for cardiac rhythm disturbances [1]. It has been estimated that 2.2 million people in America and 4.5 million in the European Union have AFIB [2]. AFIB is also a very common arrhythmia in patients who have undergone cardiac surgery and it is estimated that almost 1 in 5 patients admitted to an intensive care unit will develop atrial fibrillation [3]. Experts from the American Heart Association recommend continuous monitoring for patients at high risk for developing postoperative AFIB until hospital discharge [3].

Given the importance and prevalence of AFIB, a reliable AFIB detection algorithm is a valuable feature for ECG monitoring devices. Such an algorithm should be able to detect episodes of AFIB accurately and at the same time have a low computational complexity in order to analyze ECG signals in real time. AFIB detection algorithms usually exploit the irregularity of RR intervals and absence of the normal P wave as the most important characteristics of AFIB in the ECG signal. Several methods have been reported for AFIB detection, most of them based only on RR interval irregularity [4–7]. The irregularity measure was obtained by simple methods like variance of selected RR intervals [7] or more sophisticated methods such as Markov modeling [4], Neural Networks [5] and Hidden Markov Models [6]. Simple measures often try to capture and quantify the randomness of RR intervals while modeling approaches try to construct a model for RR irregularity. Given the chaotic nature of AFIB it is unlikely to model the exact behavior of RR irregularity during AFIB. However the models are usually helpful in distinguishing the RR irregularity of AFIB from that of other cardiac arrhythmias.

In this paper we compare two AFIB detection methods based on RR interval variability in a statistical framework. The difference between consecutive RR intervals, normalized by the mean of RR intervals, is considered random variable. We use the NP detection approach [10] to obtain criteria for AFIB detection assuming Laplace and Gaussian distribution functions for the data. Since both detection criteria are sensitive to outliers, we add a constraint on the absolute value of the RR differences and exclude those greater than a threshold. We use three databases to evaluate and compare the performance of these methods: MIT-BIH Arrhythmia database [8], MIT-BIH AFIB database [8], and an AFIB database developed in collaboration by Draeger Medical and the Thorax center, Erasmus Medical Center Rotterdam, The Netherlands [9]. The performance of the two methods are compared using ROC's over the different databases.

2. Methods

We use the NP decision approach to detect AFIB episodes from non-AFIB episodes. The random variable x for the i'th heart beat is defined as $x_i = (RR_i - RR_{i-1})/\overline{RR}_i$, where RR represents the R to R interval and $\overline{RR}$ is the mean of RR interval obtained as follows: $\overline{RR}_i = 0.9 * \overline{RR}_{i-1} + 0.1 * RR_i$.

We use the MIT-BIH Arrhythmia database as our train-

ISSN 0276–6574

ing set and calculate x for all the Normal beats surrounded by two Normal beats in the database. The histogram of x for AFIB and non-AFIB episodes are shown in Fig. 1 and Fig. 2. Based on the histograms we assume two zero mean pdf's for the data: Gaussian (P_G) and Laplace (P_L).

$$P_G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \tag{1}$$

$$P_L(x) = \frac{1}{2b} e^{-|x|/b} \tag{2}$$

The scaled representation of Gaussian and Laplace pdf's are shown along the histograms in Fig. 1 and Fig. 2. The parameters σ, b and the scale values were chosen so that the pdf's visually fit the histograms.

The NP detection criterion for X (a sequence of x with length N) is

$$L(x) = \frac{p(X|\text{AFIB})}{p(X|\text{NOAFIB})} = \frac{\prod_{i=1}^{N} p(x_i|\text{AFIB})}{\prod_{i=1}^{N} p(x_i|\text{NOAFIB})} > \gamma, \tag{3}$$

where γ is the detection threshold. It can be shown that the NP detection criterion for the case of Gaussian pdf assumption is reduced to the test of variance [10].

$$L_1(x) = \sum_{i=1}^{N} x_i^2 > \gamma_1 \tag{4}$$

It can also be shown that the NP detector in the case of Laplace pdf assumption is reduced to the test of absolute deviation.

$$L_2(x) = \sum_{i=1}^{N} |x_i| > \gamma_2 \tag{5}$$

The mean of x is assumed to be zero in (4) and (5). γ_1 and γ_2 are the detection thresholds and related to γ and the parameters of AFIB and non-AFIB pdf functions. These threshold values can be obtained by choosing a value for γ and estimating the parameters of the pdf functions or alternatively according to the ROC's over different databases as explained later.

Both variance and absolute deviation are sensitive to outliers, which are mostly due to the measurement noise and non-AFIB rhythms. To reduce the effect of outliers we exclude interval differences larger than a fixed threshold ($|RR_i - RR_{i-1}| > \beta$) from the calculation of the variance and absolute deviation.

3. Evaluation method

We used three databases to evaluate the performance of the two methods explained in Sec. 2: MIT-BIH Arrhythmia database [8], MIT-BIH AFIB database [8], and Draeger Medical AFIB database [9]. The MIT databases mostly represent paroxysmal AFIB cases while Draeger Medical database contains chronic AFIB cases. The beat annotations for MIT-BIH AFIB and Draeger AFIB databases were generated by Draeger's ECG Monitoring Algorithm which classifies the beats as Normal or Ventricular. We used reference beat annotations for MIT-BIH Arrhythmia database provided by Physionet [8].

To measure RR intervals for MIT-BIH AFIB and Draeger AFIB databases, digitized ECG signals were first processed to detect the R-waves. The signals were filtered to remove extraneous low- and high-frequency noise. Then the leads were further filtered and rectified to emphasize R-

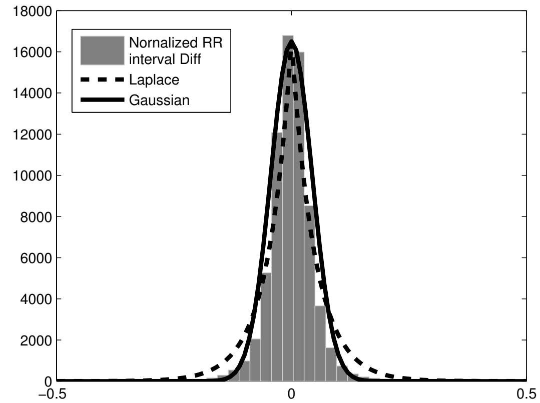

Figure 1. Histogram of the normalized RR interval difference (x) of AFIB episodes of the MIT-BIH Arrhythmia database along with scaled Gaussian and Laplace pdf's to fit the Histogram.

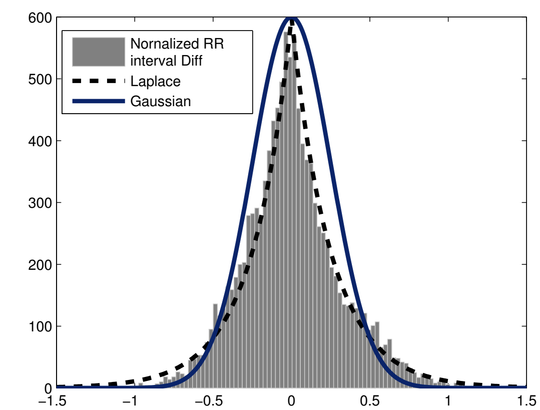

Figure 2. Histogram of the normalized RR interval difference (x) of non-AFIB episodes of the MIT-BIH Arrhythmia database along with scaled Gaussian and Laplace pdf's to fit the Histogram.

waves with respect to P- and T-waves. If two leads were available, they were combined at this stage and an adaptive threshold was applied to determine the R-waves. The R-waves were delineated to determine onset and offset, then the center-of-gravity was determined to yield a robust fiducial which is independent of R-wave polarity and morphology. RR interval was simply calculated as the time difference between sequential fiducials.

Records of the three databases were divided into frames each containing 30 heart beats. For each frame we excluded the intervals bounded by Ventricular Beats and calculated the L_1 and L_2 criteria. We also applied the interval constraints explained in Sec. 2 and calculated the L_1 and L_2 again. The interval difference threshold β was set to 300ms. This value was selected experimentally in order to obtain optimum performance.

We compared performance of the two methods in terms of duration Sensitivity (SE) and duration Positive Predictivity (PP) as explained in [9]. We calculated SE and PP for a range of values for γ_1 and γ_2 and plotted ROC curves for each method with and without interval constraints. Finally we picked the threshold values that resulted in a reasonable trade off between SE and PP for AFIB detection algorithm.

4. Results

Fig. 3 shows the ROC's corresponding to L_1 and L_2 detection criteria applied to the MIT-BIH Arrhythmia Database with and without the interval constraints. MIT-BIH Arrhythmia Database contains a variety of rare events other than AFIB that may involve RR interval irregularity, therefore it is helpful to evaluate PP of AFIB detection methods. Fig. 3 shows that AFIB detection based on L_2 performs better than the one based on L_1 before and after applying the interval constraint. Also the interval constraint improves the performance of both methods considerably.

Fig. 4 and Fig. 5 show the ROC's corresponding to L_1 and L_2 detection criteria applied to the MIT-BIH AFIB and Draeger Medical AFIB Databases with and without the interval constraints. These databases have a variety of paroxysmal and chronic AFIB cases which are helpful to evaluate the sensitivity of AFIB detection methods. However the PP over these databases may not represent the real practice statistics because they do not contain the variety of events like those in the MIT-BIH Arrhythmia Database. The performance of L_2 is better than L_1 in both cases and the interval constraint improves the performance although not as much as that of seen for MIT Arrhythmia database.

Table 1 shows the performance of AFIB detection methods based on L_1 and L_2 with interval constraints for the three different databases used in this study. The values of γ_1 and γ_2 were set to 0.016 and 0.095 respectively to result in a reasonable trade off between SE and PP on all databases. The corresponding points on the ROC's are marked with '*' on the figures.

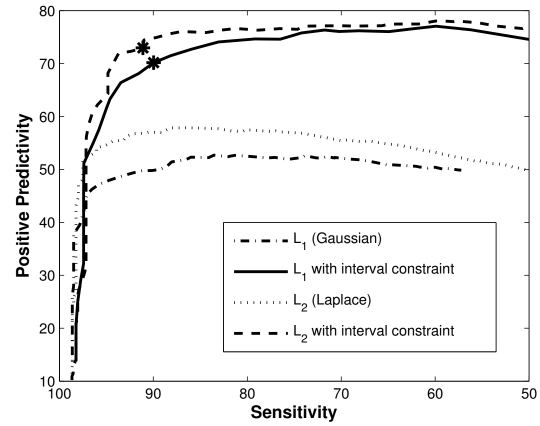

Figure 3. ROC's corresponding to L_1 and L_2 detection criteria applied to the MIT-BIH Arrhythmia Database with and without the interval constraint

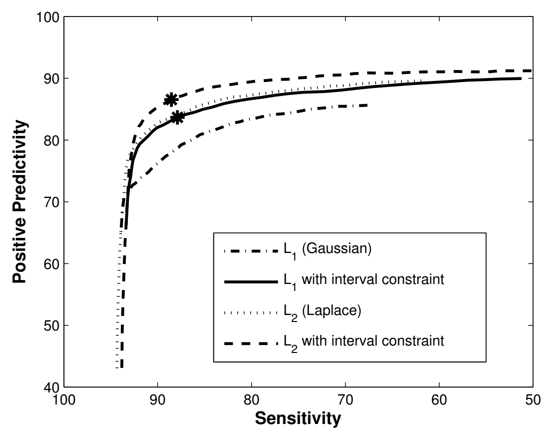

Figure 4. ROC's corresponding to L_1 and L_2 detection criteria applied to the MIT-BIH AFIB Database with and without the interval constraints

5. Discussion and conclusions

The results showed that AFIB detection based on L_2 resulted in a better performance than the one based on L_1. This could be explained by better fitness of the Laplace pdf to the data during AFIB episodes (see Fig.1). Also despite the simplicity of the proposed detection methods, the performance is comparable to more complex methods (SE=%94, PP=%86, for MIT AFIB database [4]).

However the databases used in this study are not representative of the patient population as a whole. Each of the

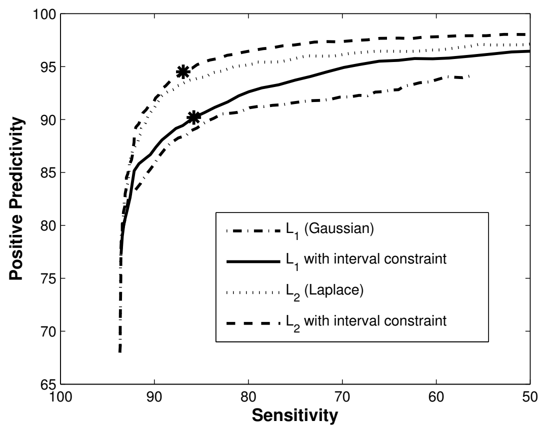

Figure 5. ROC's corresponding to L_1 and L_2 detection criteria applied to the Draeger Med. AFIB Database with and without the interval constraints

	L_1 (Gaussian) with interval constraint		L_2 (Laplace) with interval constraint	
	SE	PP	SE	PP
MIT Arr. DB	%90	%70	%92	%73
MIT AFIB DB	%88	%84	%89	%87
Drager AFIB DB	%86	%90	%87	%94

Table 1. Performance of AFIB detection methods

databases has a different emphasis and represents a specific group of patients. Nevertheless the performance of the methods against the three databases provide a great insight about the performance in different situations.

The threshold values γ_1 and γ_2 are chosen to obtain a reasonable trade off between SE and PP over the three databases simultaneously. The corresponding points on the ROC's are close to the corner for all databases which shows the detection methods are not biased to the training dataset. We note that it is not possible to determine one optimum threshold value and different clinical situations may require more sensitivity or less false alarms.

The results presented in this work are based on applying AFIB detection methods on frames of ECG signal containing 30 heart beats and there is no overlap between frames. Therefore the boundaries of the AFIB episodes can not be accurately determined. This may have a negative effect in the performance statistics but it is not significant in clinical practice. This error has a larger impact on the performance statistics of the MIT-BIH Arrhythmia database which has many AFIB episodes shorter than 10 seconds.

AFIB detection methods that are only based on RR irregularity have certain limitations. There are other rhythms that involve RR irregularities which may trigger these AFIB detection methods. We excluded Ventricular beats and added constraints on the RR intervals that were fed to the detection methods to decrease the number of false detections, however these constraints eliminated only part of the false detections.

Acknowledgements

This work was supported by Draeger Medical Systems, Inc., Andover, MA USA. The author is grateful to Teus van Dam CCRN and the Thorax Center of Erasmus Medical Center for their collaboration and support in developing the AFIB database.

References

[1] Fuster V, Ryden LE, Cannom DS et al. ACC/AHA/ESC 2006 Guidelines for the Management of Patients With Atrial Fibrillation. Circulation 2006;114:700–752.

[2] Feinberg WM, Cornell ES, Nightingale SD et al. Relationship between prothrombin activation fragment F1.2 and international normalized ratio in patients with atrial fibrillation. Stroke Prevention in Atrial Fibrillation Investigators. Stroke 1997;28:1101–1106.

[3] Drew BJ, Califf RM, Funk M et al. Practice Standards for Electrocardiographic Monitoring in Hospital Settings. Circulation 2004;110:2721–2746.

[4] Moody G, Mark R. A new method for detection atrial fibrillation using R-R intervals. Computers in Cardiology 1983;227–230.

[5] Artis S, Mark R, Moody G. Detection of atrial fibrillation using artificial neural networks. Computers in Cardiology 1991;173–176.

[6] Young B, Brodnick D, Spaulding R. A comparative study of a hidden Markov model detector for atrial fibrillation. Proceedings of the 1999 IEEE Signal Processing Society Workshop 1999;468–476.

[7] Logan B, Healey J, Robust detection of atrial fibrillation for a long term telemonitoring system. Computers in Cardiology 2005;619–622.

[8] Moody G, Mark R. The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it. Computers in Cardiology 1990;185–188.

[9] Ghodrati A, Murray WJ, Marinello S. RR interval analysis for detection of atrial fibrillation in ECG monitors. Proceedings of 30th International Conference of EMBS 2008

[10] Kay SM. Fundamentals Of Statistical Signal Processing: Detection Theory. Upper Saddle River, NJ: Prentice Hall, 1998

Address for correspondence:

Alireza Ghodrati
6 Tech Dr., Andover, MA 01810, USA
alireza.ghodrati@draeger.com