A Pattern-Recognition Approach for Lead-Selection in Heartbeat Detection

Mariano Llamedo, Juan Pablo Martínez, Pablo Laguna

BSICoS, Aragon Inst of Eng Research, IIS Aragón, Univ of Zaragoza, Aragon, Spain
CIBER of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Spain

Abstract

In this work, we developed and evaluated an algorithm for selecting the most suitable lead for performing heartbeat detection in ECG signals. For the development and evaluation we used a public dataset of 927 multilead (2-12 leads) stress-test recordings, with manually reviewed heartbeat locations. The algorithm consists of a pattern-recognition block based on features calculated from the RR interval sequence, and a mixture of Gaussian classifier. This block estimates whether the heartbeat is correctly detected, ommited or incorrectly detected. With these estimations, a detection quality index is calculated from the sensitivity (S) and positive predictive value (P^+) of each lead. With this quality index a decision is made to choose the best lead. Results show that the correct lead has been selected in 70% of the recordings, and in 93% of the recordings the best lead was among the top 3 leads with higher detection quality index. Finally, the selection of the lead with higher quality index produces a gross median S of 100%, with percentile 5 at 99.6, and a gross median P^+ of 98.9%, with percentile 5 at 89.2. The algorithm was developed and evaluated using ECG signals, but could be used with other cardiovascular signals as well, being suitable for automatically selecting the best lead/signal, or sorting them for further analysis or manual correction.

1. Introduction

The analysis of signals of cardiovascular origin such as the electrocardiographic (ECG) or plethysmographic (PPG) signals provides a noninvasive and inexpensive technique to analyze the heart function for different cardiac conditions. In addition, complementary information about the cardiovascular system can be gathered with other signals, such as the arterial blood pressure (BP). In the last years many algorithms were presented for the automatic analysis of these signals, some of them are available in Physionet [1]. One of the most frequent analysis performed in first place is the detection of heartbeats, and the subsequent construction of the RR interval sequence.

The automatic detection of heartbeats in the ECG was abundantly studied in the last decades, developing many algorithms which obtained average sensitivities (S) and positive predictive values (P^+) above the 90% in public databases [2]. These results were presented in this same conference by the same authors. We have shown [3] that the performance of the detector in multilead signals is very dependent on the chosen lead, and usually ranges from 90% to almost 100% for both S and P^+. Despite of the good performance and generalization of current algorithms, applications needing to track specific signal features in a beat-to-beat fashion rely heavily of a precise location of the heartbeats. Stress testing recordings for ischemia detection are one example where several ECG features must be tracked as the test evolves.

The objective of this work is to develop an algorithm to quantify the quality of the fiducial points produced by QRS detection algorithms. The objective of this quality metric is 1) to rank the leads in order to select the best performing lead in an unassisted operation, and 2) to serve as input to subsequent semiautomatic correction algorithm in an assisted mode.

2. Material and methods

In this work we used two ECG databases for training and evaluating the generalization of the algorithm. The first is called “Exercise testing and perfusion imaging” and is hosted by the THEW project [4]. This database includes 909 patients referred for stress testing following Bruce protocol. During the test, 12-lead ECG was recorded until the recovery phase at a sampling rate of 1000 Hz, 0.15 μV of resolution. The other database is the MIT-BIH ST Change Database, available in Physionet [1]. This database includes 18 two-lead and 10 single-lead stress-test recordings sampled at 360 Hz. Ten recordings of this database were discarded as they are single-lead. The dataset of 927 recordings which comprises both databases was divided into a train and evaluation datasets. The train dataset includes the first 20 recordings from the THEW database, while the remaining recordings were exclusively used for performance evaluation.

The algorithm was evaluated using the ECG QRS detection algorithm based on the wavelet transform described in [5], but can be adapted to any algorithm that detects heartbeats in any cardiovascular (CV) signal. The pattern-recognition approach consists in training an statistical model based on 5 features calculated from the RR interval sequence and a mixture of Gaussians classifier. This classifier will label each heartbeat as true positive (TP), false positive (FP) or false negative (FN) heartbeat. For the training of the model, the missed heartbeats were added to the dataset to model the FN class, because they are in fact missed and no present in the measurements. That is, when the algorithm is operating, it has no information from their existence other than the TP or FP detection that follows the FN. To deal with this situation the next heartbeat, either a TP or FP beat, is relabeled as FN as is shown for the third heartbeat of Figure 2.

Four of the five features used for describing the rhythm evolution, were already used in the context of heartbeats classification [6]. For the i-th QRS complex fiducial point $f_{QRS}[n]$, the corresponding RR interval is defined as $RR[n] = f_{QRS}[n] - f_{QRS}[n-1]$. Then we used $RR[n-1]$, $RR[n]$ to describe the recent evolution of the heart rhythm, and two estimates of the local and global rhythm by calculating the mean RR interval in the last 10 and 60 seconds, referred as $RR_{10}$and RR_{60} respectively. Figure 1 shows an example of the rhythm features described.

The last feature used depends on the co-occurrence ($O[n]$) of a heartbeat in other ECG leads (or CV signals). Thus, for a multilead ECG signal with N leads, a heartbeat $Q_l[n]$ detected in lead l would have a $O[n] = N$ if this heartbeat was also detected in all the other leads. Note that erroneously detected heartbeats or FP, are likely to have $O[n]$ values close to zero in multilead signals, as is shown in Figure 2. In order to use the feature $O[n]$ with an arbitrary amount of signals N, we mapped $O[n]$ through a sigmoid in order to convert the range from $[0-N]$ to an arbitrary range [0-1000]. For considering co-occurrence across leads, we seek the existence of heartbeats in other leads within a time window of 200 ms centered in $Q_l[n]$.

The classifier used is based on the mixture of Gaussians model [7], because of the expected multi-modality of the data. The training of this classifier consists of estimating the parameters of a density function

$$p(\mathbf{x}|\mathbf{\Psi}) = \sum_{k=1}^{K} \pi_k \cdot f(\mathbf{x}; \boldsymbol{\mu}_\mathbf{k}, \mathbf{\Sigma}_\mathrm{k}),$$

being

$$f(\mathbf{x}; \boldsymbol{\mu}_\mathbf{k}, \mathbf{\Sigma}_\mathrm{k}) = \frac{1}{\sqrt{(2\pi)^\mathrm{m} \, |\mathbf{\Sigma}_\mathrm{k}|}} \cdot \exp^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_\mathbf{k})^\mathrm{T} \mathbf{\Sigma}_k^{-1} (\mathbf{x}-\boldsymbol{\mu}_\mathbf{k})}$$

where the m-dimensional feature vector

$$\mathbf{x_n} = (RR[n-1],\ RR[n],\ RR_{10},\ RR_{60},\ O[n])$$

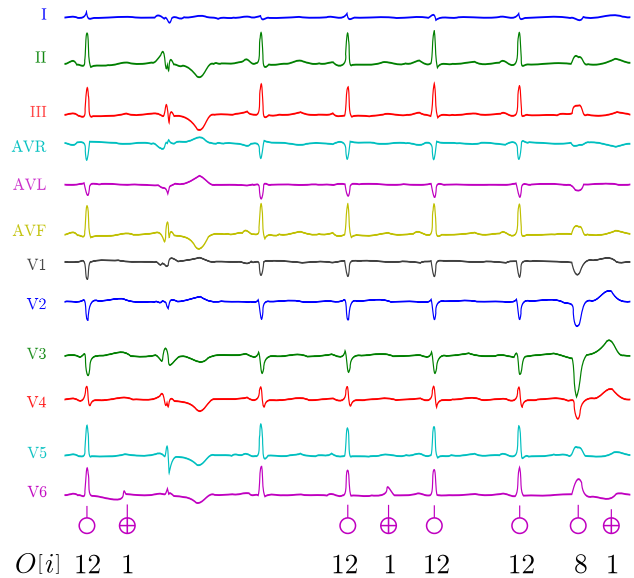

Figure 2. Toy example showing the co-occurrence feature $O[n]$, calculated for lead V6 assuming all other leads are perfectly detected. Note that false positives are showed with a +, and false negative labels are assigned to the next detection after a FN.

is modeled by K Gaussians with mixing coefficients π_k, in order to retain a more realistic structure of the data. The parameter set $\Psi = \{\pi_k, \boldsymbol{\mu}_k, \mathbf{\Sigma}_k | k = 1, \ldots, K\}$ is estimated by maximum likelihood criterion. We maximize the log likelihood

$$L(X|\Psi) = \ln \prod_{n=1}^{N} p(\mathbf{x_n}|\mathbf{\Psi}),$$

for the N heartbeats in each recording $X = \{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$. Since there is no closed-form solution for Ψ by maximizing $L(X|\Psi)$, the well-known expectation-maximization algorithm (EM) is used to obtain the estimation equations of the parameters Ψ, which are the mixing coefficient for each cluster

$$\hat{\pi}_k = \frac{1}{N} \sum_{m=1}^{N} \hat{\beta}_{m,k}, \tag{1}$$

the cluster mean

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{N\hat{\pi}_k} \sum_{m=1}^{N} \hat{\beta}_{m,k} \mathbf{x_m} \tag{2}$$

and cluster covariance matrix

$$\hat{\mathbf{\Sigma}}_\mathrm{k} = \frac{1}{\mathrm{N}\hat{\pi}_\mathrm{k}} \sum_{\mathrm{m}=1}^{\mathrm{N}} \hat{\beta}_{\mathrm{m,k}} (\mathbf{x_m} - \hat{\boldsymbol{\mu}}_\mathbf{k}) \cdot (\mathbf{x_m} - \hat{\boldsymbol{\mu}}_\mathbf{k})^\mathrm{T}. \tag{3}$$

Where $\hat{\beta}_{m,k}$ is known as the ownership variable, which

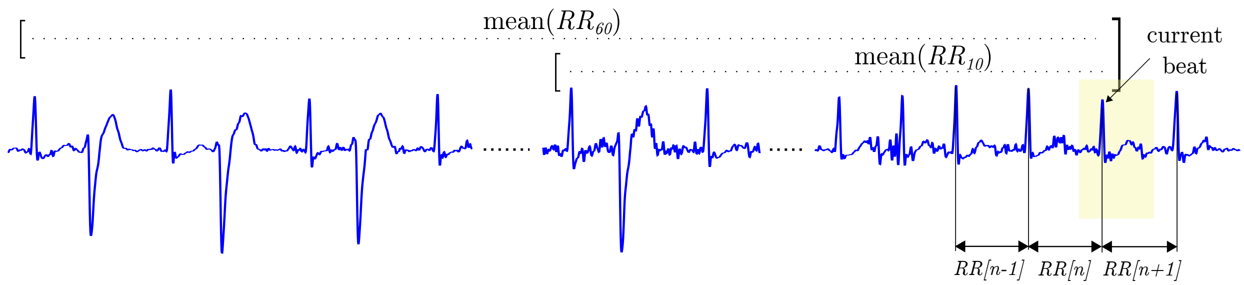

Figure 1. Features used by the algorithm based on the RR interval sequence.

indicates the probability of sample $\mathbf{x_m}$ to have been generated by the k-th component

$$\hat{\beta}_{m,k} = \frac{\hat{\pi}_k \cdot f(\mathbf{x_m}; \hat{\boldsymbol{\mu}}_\mathbf{k}, \hat{\boldsymbol{\Sigma}}_\mathrm{k})}{\sum_{j=1}^{K} \hat{\pi}_j \cdot f(\mathbf{x_m}; \hat{\boldsymbol{\mu}}_\mathbf{j}, \hat{\boldsymbol{\Sigma}}_\mathrm{j})}.$$

The EM algorithm iteratively computes the weight, location and dispersion for each of the K clusters (Eq. (1)-(3) respectively), until $\hat{\beta}_{m,k}$ does not change significantly, which is equivalent to obtaining stable clusters. The interested reader is referred to [7, 8] for details, equations and the implementation used. Then we learn in the training data a parameter set Ψ_i for each of the three classes, that are used in the evaluation data to calculate three posterior probabilities for each heartbeat. Finally during the use of the trained classifier, each heartbeat is assigned to the class with higher posterior probability.

With the estimated and true labels we construct a 3×3 confusion matrix to calculate the class and global performances:

$$\text{Estimated classes}$$

$$\begin{array}{c c c c}
 & \begin{array}{ccccc} 1 & \dots & i & \dots & C \end{array} & \\
\text{True classes} \begin{array}{c} 1 \\ \\ i \\ \vdots \\ C \end{array} & \begin{pmatrix} n_{11}^T & \dots & n_{1i}^F & \dots & n_{1C}^F \\ \vdots & \ddots & \vdots & & \\ n_{i1}^F & \dots & n_{ii}^T & \dots & n_{iC}^F \\ \vdots & & \vdots & \ddots & \\ n_{C1}^F & \dots & n_{Ci}^F & \dots & n_{CC}^T \end{pmatrix} & \begin{array}{c} N_1 \\ \vdots \\ N_i \\ \vdots \\ N_C \end{array} \\
 & \begin{array}{ccccc} P_1 & \dots & P_i & \dots & P_C \end{array} & N_T.
\end{array}$$

For the i-th class n_{ii}^T is the number of correctly classified examples and n_{ij}^F is the number of examples of class i classified as class j; N_i is the total number of examples for class i, P_i is the number of examples classified as class i and N_T is the total number of heartbeats in the dataset. Being $N_i = n_{ii}^T + \sum_{m \neq i} n_{im}^F$, $P_i = n_{ii}^T + \sum_{m \neq i} n_{mi}^F$, and $N_T = \sum_{i=1}^{C} N_i$, then for each class we define class sensitivity $S_i = \frac{n_{ii}^T}{N_i}$ and class positive predictive value $P_i^+ = \frac{n_{ii}^T}{P_i}$ as and and its global counterparts as the mean of all class performances $S = \frac{1}{C}\sum_{i=1}^{C} S_i$ and $P^+ = \frac{1}{C}\sum_{i=1}^{C} P_i^+$. Finally the global accuracy is calculated as

$$A = \frac{1}{N_T}\sum_{i=1}^{C} n_{ii}^T = \sum_{i=1}^{C} \frac{N_i}{N_T} S_i \tag{4}$$

Using the estimated labels we finally calculated a quality index

$$q_l = {}^2\!/_3 \hat{S}_l + {}^1\!/_3 \hat{P_l^+}$$

for each lead l where $\hat{S}_l = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$ and $\hat{P_l^+} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$. Then for each recording, we choose the lead with higher q_l for performance calculation.

3. Results

The results obtained in the evaluation set are presented in Tables 1 and 2. As it can be seen, the S and P^+ obtained for FP and TP classes are above 80%, however the P^+ of FN class shows that there is room of improvement. Despite this weakness, the calculated quality metric q_l is robust enough to choose the correct lead in 70% of the evaluated recordings and in the 93% the lead chosen was among the 3 best leads. When choosing the lead with highest quality index, the median decrease in S, compared to the actual best lead was 0% with percentile 5 at 0.11% and percentile 95 at -0.05%, while for the P^+ the median decrease of 0%, with same percentiles at 0.49 and 0% respectively.

4. Discussion and conclusions

In this work we presented an algorithm for selecting the best single-lead QRS detection from an arbitrary multilead signal. This methodology was evaluated for a QRS detector developed in our group [5] and for stress-testing recordings, but could be applied to any detector and other ECG recording types. One aspect not studied in this work is the minimum window length necessary for the algorithm to produce a useful quality estimation q_l. When used in shorter segments, this analysis would allow not only to select the best lead in shorter recordings, but also the best

Table 1. Classification results obtained in evaluation set.

		Algorithm fn	fp	tp	Total
Truth	FN	313319	2175	27714	343208
	FP	104054	1096893	76190	1277137
	TP	2099566	231492	17332305	19663363
	Total	2516939	1330560	17436209	21283708

FN		FP		TP		Total		
S	P^+	S	P^+	S	P^+	A	S	P^+
91	12	86	82	88	99	88	88	65

Table 2. Gross QRS detection performance. Median and percentiles 5 and 95.

lead selected	S		P^+	
best	100	(99.7 – 100)	99	(89.5 – 100)
this algorithm	100	(99.6 – 100)	98.9	(89.2 – 100)
worst	93.2	(76.4 – 99.9)	89.1	(69.5 – 98.9)

lead for each segment in larger recordings, in order to create a combined RR interval series from the concatenation of the different segments.

Most of the features were previously validated in the context of heartbeat classification, as well as the classifier used [6, 9]. The only feature newly-proposed in this work is the co-occurrence among the remaining leads/signals, which exploits the multilead/multisignals nature of cardiovascular recordings.

In order to perform a realistic evaluation of the performance, the algorithm was trained only in 20 registers, and evaluated in the remaining 907. The results suggest that the classifier part of the algorithm performed well above the 80% labeling the FP and TP classes, as can be seen in Table 1. The most difficult class was undoubtedly FN, with a good sensitivity, but a very low P^+. This low performance is influenced by the high sensitivity of the QRS detector used, in this case the detector presented in [5], producing very few FN heartbeats to train the system. As can be seen in Table 1, the FN class is 4 times less represented than FP, and 60 times less represented than TP. However this lack of performance seems not to be preponderant in the total performance of the algorithm, since the correct lead was chosen in 70% of the recordings, and was one of the three leads with higher q_l index in the 93% of the recordings. The last result also suggest that the algorithm is suitable for selecting the correct lead, since the median decrease in S and P^+ was 0%. That suggest that even if the lead was not the best in one given recording, the performance is comparably good. It is also remarkable the improvement obtained respect to the worst lead, as can be seen in table 2.

The work presented can be extended to other classes of ECG recordings and cardiovascular signals, such as BP and PPG, and also for shorter time windows in order to deal with transient signal loss. The results presented suggest that the algorithm performance is suitable for its use in ECG stress-test recordings.

Acknowledgments

This work was supported by projects TEC2010-21703-C03-02, TEC2013-42140-R from MINECO (Spain) and by Grupo Consolidado BSICoS (T96) from DGA (Aragon) and European Social Fund (EU). The CIBER of Bioengineering, Biomaterials and Nanomedicine is an initiative of ISCIII.

References

[1] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[2] Elgendi M. Fast qrs detection with an optimized knowledge-based method: Evaluation on 11 standard ecg databases. PLoS ONE 09 2013;8(9):e73557. URL `http://dx.doi.org/10.1371/journal.pone.0073557`.

[3] Llamedo M, Martínez JP. Qrs detectors performance comparison in public databases. In Computers in Cardiology 2014, volume 41. IEEE Computer Society Press. (Submitted), 2014; .

[4] Couderc JP. The telemetric and holter ECG warehouse initiative (thew). URL `thew-project.org`.

[5] Martínez JP, Almeida R, Olmos S, Rocha A, Laguna P. A wavelet-based ECG delineator: Evaluation on standard databases. IEEE Transactions on Biomedical Engineering 2004;51:570–581.

[6] Llamedo M, Martínez JP. Heartbeat classification using feature selection driven by database generalization criteria. IEEE Transactions on Biomedical Engineering march 2011; 58(3):616–625. ISSN 0018-9294.

[7] van der Heijden F, Duin R, de Ridder D, Tax D. Classification, Parameter Estimation and State Estimation. John Wiley & Sons, 2005.

[8] Duin R, Juszczak P, Paclik P, Pekalska E, deRidder D, Tax D, Verzakov S. PR-tools, a matlab toolbox for pattern recognition, 2008. URL `http://www.prtools.org`.

[9] Llamedo M, Martínez JP. An automatic patient-adapted ecg heartbeat classifier allowing expert assistance. Biomedical Engineering IEEE Transactions on aug. 2012;59(8):2312 – 2320. ISSN 0018-9294.

Address for correspondence:

Mariano Llamedo Soria, llamedom@unizar.es
I3A - I+D Building, C/ Mariano Esquillor S/N
Despacho 4.0.05 – CP: 50018, Zaragoza, España.