31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Unsupervised Feature Selection in Cardiac Arrhythmias Analysis

J. L. Rodríguez-Sotelo · D. Cuesta-Frau,
D. Peluffo-Ordóñez · G. Castellanos-Domínguez

Abstract— **The problem of detecting clinical events related to cardiac arrhythmias in long term electrocardiograms is a difficult one due to the large amount of irrelevant information that hides such events.**

This problem has been addressed in the literature by means of clustering or classification algorithms that create data partitions according to a cost function based on heartbeat features dissimilarity measures.

However, studies about the type or number of heartbeat features is lacking. Usually, the feature sets used are relevant but redundant, which degrades algorithm performance.

This paper describes a method for automatic selection of heartbeat features. This method is assessed using real signals from the MIT database and common features used in previous works.

I. INTRODUCTION

With regard to cardiac arrhythmias, heart condition is assessed in terms of rate, regularity, impulse origin or conduction, and waveforms [16], commonly measured by means of an electrocardiogram (ECG).

However, since an ECG is a nonstationary signal, the significant clinical events may be randomly scattered in time throughout the entire signal. Additionally, these events might be very infrequent. Consequently, to assure no missing arrhythmia information, ECGs are recorded for hours or even days. Long term ECGs involve a vast amount of data, most of it not clinically significant, which complicates its analysis both in terms of computational time and event detection capabilities.

Holter records are the usual ECG type to seek possible arrhythmias [13]. Specific signal features correlated with arrhythmias are analized. These features can be, in general, waveform morphology, amplitude, or time or frequency related (RR interval, heart rate) [4], [14], [15]. Sometimes the features are corrupted by signal artifacts [18] and by the intrinsic variability of heartbeats [8]. Therefore, it is absolutely necessary to develop efficient and accurate algorithms able to deal in a short time with the data, while not omitting relevant clinical details. Thus, new computer tools could be developed as diagnosis aid systems [3].

These algorithms or methods will have to include a feature selection stage. Then, heartbeats are classified using a supervised or an unsupervised approach. A number of methods have been reported in the scientific literature to carry out this classification. For instance, [6], [3] and [11] using a supervised approach, or [9], [4] and [12] using an unsupervised one. However, little attention has been paid to the feature selection stage, specially regarding the significance of the features used. Such study could improve the performance of the methods by focusing on the most relevant information in data.

This work describes a scheme for unsupervised ECG feature selection based on a two stage recurrent algorithm. It algebraically selects the most relevant features, computed over the heartbeats affinity matrix [17]. The experimental set used was drawn from the MIT-BIH database [10], including normal (N) heartbeats, as well as the arrhythmia types recommended by the AAMI [1], usually found in Holter records: ventricular extrasystoles (VE), left and right branch bundles (LBBB and RBBB) and atrial premature beats (APB). A set of features are ranked by means of a non-supervised scheme based on the gaussian maximization clustering algorithm (GEM) [2]. Cluster initialization is carried out using a J-means [7] algorithm. The results were assessed in terms of sensitivity and specificity measures, based on the heartbeat labels.

II. MATERIALS AND METHODS

A. Feature Set

As the initial feature set, we chose features taken from previous works that achieved good performance in wave morphology characterization, signal variability, and signal representation. They have been employed in applications to detect N, VE [4], LBBB [5], RBBB [9] and APB [13]. Table I shows, in detail, the feature set, including its description and computation, if applicable.

B. Feature Relevance Measurement

To assess the relevance of each feature, we used the algorithm described in [17]. In proceeds as follows:

Definition 1: **(Relevant Features Optimization)** Let M be an input matrix of size $n \times q$, where n accounts for the features and q for the observations. The matrix rows are $\mathbf{m}_1^\top, \ldots, \mathbf{m}_n^\top$. Let $A_\alpha = \sum_{i=1}^{n} \alpha_i \mathbf{m}_i \mathbf{m}_j^\top$, for some unknown scalars $\alpha_1, \ldots, \alpha_n$. The weight vector $\alpha = (\alpha_1, \ldots, \alpha_n)^\top$ and the matrix Q of size $q \times k$, are determined at the maximal

This work was supported by GCPDS and GIICA research groups

J. L. Rodríguez-Sotelo, D. Peluffo-Ordóñez and G. Castellanos-Domínguez are with Faculty of Electrical and Electronic Engineering, Universidad Nacional de Colombia sede Manizales, Colombia jlrodriguezso@unal.edu.co

D. Cuesta is with the Technological Institute of Informatics, Polytechnic University of Valencia-Alcoi Campus, Alcoi, Spain, dcuesta@disca.upv.es

TABLE I
FEATURE SET USED IN ARRHYTHMIA ANALYSIS WHOSE RELEVANCE HAS BEEN STUDIED.

Index	Type	Feature description
1	HRV [13]	• RR interval (r_1)
2		• pre-RR interval (r_0)
3		• post-RR interval (r_2)
4	Prematurity [13]	• RR and pre-RR intervals difference, $f_4 = r_1 - r_0$
5		• Post-RR and RR intervals difference, $f_5 = r_2 - r_1$
6		• continuous A beats ◦ $f_6 = \left(\frac{r2}{r1}\right)^2 + \left(\frac{r0}{r1}\right)^2 - \left(f_4 + \beta \cdot \frac{1}{3}\sum_{i=0}^{2} r_i^2 \cdot \log(r_i)^2\right)$, $0 < \beta < 0.1$
7	Morphology [13], [4]	• Dynamic time warping (DTW) between P wave actual and P wave template
8		• QRS complex polarity ◦ Let b be the samples of a heartbeat, then, $f_8 = P_i = \left\lvert\frac{\max(b_i)}{\min(b_i)}\right\rvert$
9		• QRS complex energy ◦ $f_9 = E_{b_j} = \sum_{i=1}^{n} b_j(i)^2$
$10,\ldots,19$	Representation [9]	• 10 Hermite coefficients ◦ $\underset{i=10:19}{f_i} = \underset{n=0:9}{\phi_n^\sigma} = \frac{e^{-t^2/2\sigma^2}}{\sqrt{2^n n!\sqrt{\pi}}} H_n(t/\sigma)$, where H_n is Hermite polynomial y σ window width.
$20,\ldots,90$	Rrepresentation [5]	• Wavelet Db2 (A4: $20-25$, D4: $26-31$, D3: $32-41$, D2: $43-58$, D1: $59-90$) ◦ Using Discrete Wavelet Transform (DWT)

point of the optimization problem:

$$\begin{aligned} &\max_{Q,\alpha_i} trace(Q^\top A_\alpha^\top A_\alpha Q) \\ &\text{subject to } \sum_{i=1}^{n} \alpha_i^2 = 1, Q^\top Q = I \end{aligned} \tag{1}$$

An optimal solution has to be found for 1. If the weight vector α is known, then the solution for the matrix Q is found by using Singular Value Decomposition (SVD) of the symmetric and positive definite matrix A_α. Otherwise, if Q is known, then α is easily determined since:

$$\begin{aligned} trace(Q^\top A_\alpha^\top A_\alpha Q) = &\sum_{i,j} \alpha_i \alpha_j (\mathbf{m}_i^\top \mathbf{m}_j) \mathbf{m}_i^\top Q Q^\top \mathbf{m}_j \\ = &\ \alpha^\top G \alpha \end{aligned} \tag{2}$$

where $G_{ij} = (\mathbf{m}_i^\top \mathbf{m}_j)\mathbf{m}_i^\top Q Q^\top \mathbf{m}_j$ is symmetric and positive definite. The optimal α is therefore the solution of the optimization problem:

$$\max_{\alpha} \alpha^\top G \alpha \text{ subject to } \alpha^\top \alpha = 1, \tag{3}$$

where the resulting α is the leading G eigenvector. Convergence to a local maximum is guaranteed by starting with an initial guess for α and iteratively compute Q given α and α computation given Q until convergence. This procedure is termed basic $Q-\alpha$.

A more advanced method, used in this study, to achieve a higher convergence rate and higher result accuracy, is to embed the α computation with the orthogonal iteration cycle to compute the k eigenvectors, as described next.

Definition 2: **(*Power-Embedded* $Q-\alpha$ Method)** Let M be an input matrix of size $n \times q$ with rows $\mathbf{m}_1^\top, \ldots, \mathbf{m}_n^\top$, and an orthonormal matrix $Q^{(0)}$ of size $q \times k$. The following steps have to be performed throughout a cycle of iterations of index $r = 1, 2, \ldots$:

Algorithm 1 Power-Embedded $Q-\alpha$ standard method

for $r = 1$ to n **do**
1. Let $G^{(r)} = \left(MM^\top\right)\left(MQ^{(r-1)}Q^{(r-1)^\top}M^\top\right)$
2. Let $\alpha^{(r)}$ be the largest $G^{(r)}$ eigenvector.
3. Let $A^{(r)} = M^\top D^{(r)} M$, where $D^{(r)}$ is $D^{(r)} = diag(\alpha^{(r)})$
4. Let $Z^{(r)} = A^{(r)} Q^{(r-1)}$
5. $Z^{(r)} \xrightarrow{QR} Q^{(r)} R^{(r)}$

end for

This method is significantly more efficient than the basic method described before (Definition 1) and achieves higher performance [17]. The steps 4 and 5 consist of the orthogonal iteration, namely, if only these steps were repeated, it would converge to the eigenvectors of $A^{(r)}$. However, the algorithm does not repeat the steps 4 and 5 only, it recomputes the weight vector α (steps 1,2 and 3) before applying another cycle of steps 4 and 5. It is demonstrated in [17] that the addition of α recalculation does not alter the convergence property of the ortogonal iteration method, and therefore the complete method converges to a local maximum.

C. Clustering

The selected feature set is clustered in two stages. The first one corresponds to centroid initialization. It is based on the

J-H-means clustering algorithm [7]. The objective function for this algorithm is the Minimum Sum-of-Squares (MSS) [13]. After a random initialization, every point p_j out of a sphere of radius ε with center at q_i, $i = 1:m$ (q_i is the i-th of the m centroids) is considered a centroid candidate. Thus, p_j replaces a current centroid q_k. After updating and computing the MMS, the process is repeated while MMS becomes more optimal. The process stops when there is no further MMS optimization. This method is described in detail in [13]. This initialization is aimed at avoiding convergence to local minimums.

The second clustering stage computes the final partition. It is based on the Gaussian Expectation Maximization Clustering (GEMC) algorithm [5]. Its objective function is a linear combination of gaussian distributions centered at each centroid:

$$GEM(X,C) = -\sum_{i=1}^{n} log\left(\sum_{j=1}^{k} p(x_i/q_j)p(q_j)\right), \qquad (4)$$

where $p(x_i/q_j)$ is the probability of x_i, since it is generated by a gaussian distribution centered at q_j, and $p(q_j)$ is the a priori probability of the cluster whose centroid is q_j. The log function is used for simplicity, and the minus sign accounts for minimization. The member and weight functions are respectively:

$$\begin{aligned} m_{GEM}(q_j/x_i) &= \frac{p(x_i/q_j)p(q_j)}{p(x_i)} \\ w_{GEM}(x_i) &= 1 \end{aligned} \qquad (5)$$

The GEMC employs a soft member function, assigning a membership level to x_i for every cluster. The Bayes rule is used to compute m_{GEM}, where $p(x_i) = \sum_{j=1}^{k} p(x_i/q_j)$. The term $p(x_i/q_j)$ can be computed as:

$$p(x_i/q_j) = f(x_i,\mu,\Sigma_j) = \frac{e^{-\frac{1}{2}(x_i-\mu)\Sigma_j^{-1}(x_i-\mu)^t}}{|\Sigma_j|^{\frac{1}{2}}(2\pi)^{-d/2}} \qquad (6)$$

where μ is the center ($\mu = q_j$), d is the size ($d = n$), and Σ is the covariance matrix. According to Bayes's rule, the matrix Σ_j can be unique ($\Sigma_j = \Sigma = cov(X)$) or compute one for each cluster ($\Sigma_j = cov(C_j)$, $j = 1,\ldots,k$).

III. Results and Discussion

A. Results

The results are shown in Figs. 2 and 1, and in Table II. Fig. 2 shows the relevance of the feature set in x axis, from f_1 to f_{90} for different records of the experimental dataset, selected randomly. The relevance falls in the interval 0 to 1. Fig. 1 depicts an example where the heartbeat set of record 207, exhibits better separability for relevance vector maximums than for minimums. Table II shows the clustering results using the selected features. The record is shown in the first column, the number and heartbeat type is shown in the second column, and the performance measures in the third columns (Sensitivity (*Se*), Specificity (*Sp*), Positive Predictive Accuracy (*PPA*) and Clustering Performance (*CP*)), described in [13]. The number of clusters for all the experiments was set to 8. It is a tradeoff between computational cost and accuracy.

TABLE II
CLUSTERING RESULTS USING THE FEATURE SET DESCRIBED IN THE TEXT

Rec.	Beat type					Performance measure%			
	N	L	R	V	A	*Se*	*Sp*	*PPA*	*CP*
207		1457	85	105	106	100	100	100	100
111		2121		1		100	100	100	100
109		2490		38		97.3	100	99.6	99.7
118			2164	16	96	100	100	100	100
212	922		1824			100	99.6	100	99.8
214		2000		256		94.9	99.3	95.8	97.3
217	244			162		99.4	99.6	99.6	99.6
105	2524			41		100	100	100	100
total	3690	8068	4073	619	202	**98.9**	**99.8**	**99.4**	**99.6**

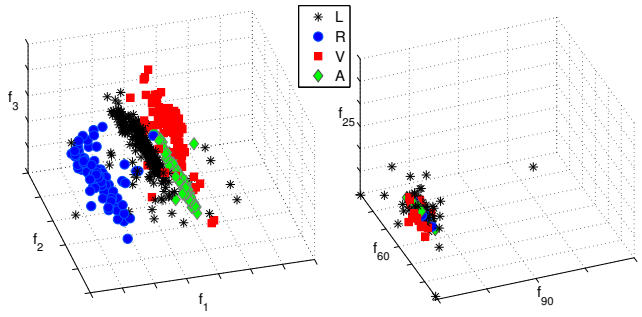

Fig. 1. Features with high relevance vs. low relevance for record 207, selected by the Q-alpha algorithm 1.

B. Discussion

As can be seen in Fig. 2, regardless of the heartbeat types in the record, the relevance analysis modifies the weight of each feature according to the inter-cluster separability. Namely, the relevance of feature subset S can be understood as the cluster coherence, using the selected features, measured by means of the afinity matrix A_s. This matrix, in a quadratic form, shows that its k larger eigenvalues can represent the cluster coherence, and the corresponding eigenvectors, the coordinate weights in such cluster. Note that the number of iterations needed (n value in the algorithm described above) for the Q-α to converge, is lower or equal to 4 in all the cases studied in this work. Therefore, the computational cost can be kept low even for long term records.

Regarding Fig. 1, it can be demonstrated experimentally that there can be feature subsets that exhibit better cluster coherence than others, although all the features used in this study have been successfully used in other works (Table I). Table II shows high performance clustering results, both $Se > 94.9\%$ and $Sp > 99.3\%$.

IV. Conclusion

This work describes a methodology to optimize the feature selection stage for the analysis of Holter records. It demonstrates that a feature subset can offer a good performance

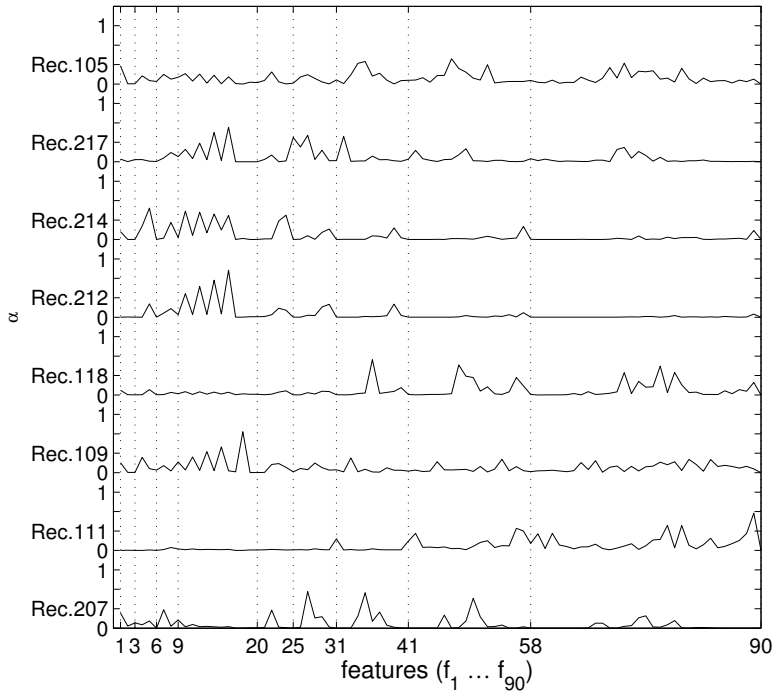

Fig. 2. Relevance vector α of the feature set (Table I) for some records from the MIT/BIH database .

from the cluster separability point of view. This method can also be applied to other feature sets.

The method proposed achieves a clustering performance higher than 98 % on average. This is very useful to analyse Holter records in an unsupervised manner. As future work, we plan to study several algorithm improvements based on the Laplacian spectrum of the afinity matrix, which adds a normalization term to it. This assures its convergence for cases where very low or very high feature values pose problems to compute the maximum eigen vector in the algorithm 1.

REFERENCES

[1] AAMI. *Recomended Practice for Testing and Reporting Performance Results of Ventricular Arrhythmia Detection Algorithms. ANSI/AAMI EC57:1998/(R)2003*. Association for the Advancement of Medical Instrumentation, Arlington, VA 22201-5762, 1999.

[2] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press Inc., New York, 1995.

[3] R. Ceylan and Y. Özbay. Comparison of fcm, pca and wt techniques for classification ecg arrhythmias using artificial neural network. *Expert Systems with Applications*, 33(2):286 – 295, 2007.

[4] D. Cuesta, M. Biagetti, R. Quinteiro, P. Mico-Tormos, and M. Aboy. Unsupervised classification of ventricular extrasystoles using bounded clustering algorithms and morphology matching. *Medical and Biological Engineering and Computing*, 45(3):229–239, 2007.

[5] D. Cvetkovic, E. D. Übeyli, and I. Cosic. Wavelet transform feature extraction from human ppg, ecg, and eeg signal responses to elf pemf exposures: A pilot study. *Digital Signal Processing*, 18(5):861 – 874, 2008.

[6] P. De-Chazal, M. O'Dwyer, and R. Reilly. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. *IEEE transaction on biomedical engineering*, 51(7):1196–1206, 2004.

[7] P. Hansen and N. Mladenovic. J-means: a new local search heuristic for minimum sum of squares clustering. *Pattern Recognition*, 34:405–413, 2001.

[8] N. Hughes, S. Roberts, and L. Tarassenko. Semi-supervised learning of probabilistic models for ecg segmentation. volume 1, pages 434–437, Sept. 2004.

[9] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sörnmo. Clustering ecg complexes using hermite functions and self-organising maps. *IEEE trans. on. Biomed.*, 48:838–847, 2000.

[10] G. Moody and M. R. The impact of the mit-bih arrhythmia database. *IEEE Transactions on engineering, medicine and biology*, 15(1):34–50, 1985.

[11] S. Osowski, L. Hoai, and T. Markiewicz. Support vector machine based expert system for reliable heart beat recognition. *IEEE trans. on. Biomed.*, 61:1–7, 2002.

[12] M. Paoletti and C. Marchesi. Discovering dangerous patterns in long-term ambulatory ecg recordings using a fast qrs detection algorithm and explorative data analysis. *Computer Methods and programs in biomedicine*, 82:20–30, 2006.

[13] J. Rodríguez-Sotelo, D. Cuesta-Frau, and G. Castellanos-Domínguez. Unsupervised classification of atrial heartbeats using a prematurity index and wave morphology features. *Medical and Biological Engineering and Computing*, 47(7):731–741, 2009.

[14] J. Sahambi, S. Tandon, and R. P. Bhatt. Using wavelet transform for ecg charecterization. an on-line digital signal processing system. *Engineering in Medicine and Biology Magazine, IEEE*, 16:77–83, january 1997.

[15] R. Silipo. Investigating electrocardiographic features in fuzzy models for cardiac arrhythmia classification. In *IDAMAP-99*. Intelligent data analysis in medicine and pharmacology, Washington, dc, October 1999.

[16] M. S. Thaler. *The only EKG book you'll ever need*. Lippincott Williams & Wilkins, 5th edition, September 2006.

[17] L. Wolf and A. Shashua. Feature selection for unsupervised and supervised inference: The emergence of sparsity in a weight-based approach. *Journal of Machine Learning Research*, 6:1855–1887, 2005.

[18] A. Ziarani and A. Konrad. A nonlinear adaptive method of elimination of power line interference in ecg signals. *IEEE transactions on biomedical engineering*, 49(6):540–547, 2002.