Lempel-Ziv Complexity and Shannon Entropy-based Support Vector Clustering of ECG Signals

Sergio Villazana*, César Seijas, Antonino Caralli

Centro de Investigación en Bioingeniería, Facultad de Ingeniería, Universidad de Carabobo, Valencia, Venezuela.

Abstract.-

This work presents a Support Vector Machine (SVM)-based clustering method to cluster normal and pathological ECG signals on a Lempel-Ziv (LZ) complexity and Shannon entropy (SE) space. One normal ECG and three ECG signals with arrhythmic processes were selected from the MIT-BIH Arrhythmia Database and, those were processed to remove muscle and breathing noise, electrode motion artifacts, power line interference and DC offset. Each ECG signal was divided into 35 four-second segments. Training Input data to the SVM-based clustering machine were obtained by applying the LZ complexity algorithm and SE to each 35 segments of ECG signals. SVC machine was capable to separate the ECG signals (each signal represents a group) in four clusters (with an accuracy of 95.7 %) according to the four different ECG records chosen for this study.

Keywords: support vector clustering, ECG signal, arrhythmia, Lempel-ZIV complexity, Shannon entropy.

Agrupación Basada en la Complejidad Lempel-Ziv y la Entropía de Shannon de Señales de ECG

Resumen.-

Este trabajo presenta un método de agrupación basado en Máquinas de Vectores de Soporte (SVM, de sus siglas en inglés Support Vector Machines) para agrupar las señales electrocardiográficas (ECG) normales y patológicas en un espacio de complejidad Lempel-Ziv y entropía de Shannon. Un registro ECG normal y tres registros ECG con procesos arrítmicos fueron seleccionados de la base de datos de arritmias del MIT-BIH, y procesados para eliminar el ruido por movimientos musculares y por respiración, artefactos debido al movimiento de los electrodos, interferencia de línea de alimentación y componentes DC. Cada señal ECG se dividió en 35 segmentos de cuatro segundos. Los datos de entrada para el entrenamiento de la máquina de agrupamiento basada en SVM fueron obtenidos de la aplicación del algoritmo complejidad LZ y ES a cada uno de los 35 segmentos de las señales de ECG. La máquina SVC fue capaz de separar las señales ECG (cada señal representa a un grupo) en cuatro grupos (con una precisión del 95,7 %) de acuerdo a los cuatro registros de ECG seleccionados para este estudio.

Palabras claves: agrupación basada en vectores de soporte, señales ECG, arritmia, complejidad Lempel-Ziv, entropía de Shannon

Recibido: Febrero 2015
Aceptado: Marzo 2015.

1. Introduction

The electrocardiography is a non-invasive procedure for recording electronically electrical changes of the heart activity. The record of the electrical activity of the heart is commonly known as an electrocardiogram (ECG). By means of the ECG analysis, a physician can detect and diagnose several heart abnormalities. Abnormal cardiac rhythms

*Autor para correspondencia
Correo-e: svillaza@uc.edu.ve (Sergio Villazana)

of the heart are known as arrhythmias, which are often detectable on an electrocardiogram. Arrhythmias often reveal stochastic changes [1] and may vary from person to person and also there are dependent on other factor such as age and patient sex. There are many studies for detection of arrhythmias [2], by mean of the use of principal component analysis [2, 3], hidden Markov models [4], cluster analysis [5], power-frequency analysis [2, 6].

Kolmorov [7] defined the complexity of a string of a sequence (of numbers, letters, etc.) as the number of bits needed to make the shortest computer program which is capable to generate that string of sequence. In general, the complexity measures the randomness grade of a sequence. Usual measures of complexity are symbolic pattern analysis and entropy. Lempel-Ziv complexity (LZC) is an example of symbolic pattern analysis. LZC of a sequence, which was defined by Lempel and Ziv [8], measures the number of different patterns in a sequence

The entropy of a random variable is defined in terms of its probability distribution, and it is a good measure of randomness or uncertainty, that is, entropy is a measure of uncertainty or order in a signal. Low entropy (close to 0) means high order or low complexity, and high entropy (close to 1) means low order or high complexity. Shannon [9] gave some answers to the question of how to measure the amount of uncertainty in classical probability theory. He established the Shannon's entropy as the way to measure the amount of uncertainty by a probability distribution function on a finite set [10, 11].

Clustering is an unsupervised classification (natural grouping) of data into groups (clusters) [12]. Unsupervised classification means clustering classifies unknown groups (without a prior labeling of classes) while supervised classification classifies known groups (each class has assigned a label which identifies that class). Clustering process needs a cluster validation, which is a fundamental task in clustering analysis [2]. Because of a validation is a subjective problem, there is no unique definition of a cluster or unique clustering result for a given data set. Support Vector Machine (SVM) is an artificial intelligent branch whose applications in engineering multidisciplinary areas are powerful. In general, SVM maps some input vectors (input data) into a high-dimensional feature space through a nonlinear mapping to find an optimal separating hyperplane that best fit the input data. This SVM-based emergent computation algorithm was developed by Vapnik and his collaborators [7]. SVM have been implemented in regression, classification and clustering applications.

Support Vector Clustering (SVC) is an algorithm which can detect arbitrary shapes of clusters from a data space [13]. SVC maps data points from data space into a high dimensional feature space using a Gaussian kernel function [14]. The SVC algorithm will look for the smallest sphere that encloses the image of the data in the feature space. This minimal sphere is defined by some images of the input data called support vectors. When the minimal sphere in feature space is mapped back into data space, it is transformed into several contours, and each data point is enclosed by one of them. Of course, each contour is defined by the support vectors in the data space. Each contour determines a cluster and, the number of clusters is controlled by the width parameter of the Gaussian function; if this parameter is decreased imply an increasing of clusters and vice versa.

This study proposes a classification process of ECG signals using as artificial intelligent tool a Support Vector Clustering (SVC) machine. The input to SVC corresponds to the LZ complexity and the Shannon entropy features of ECG signals, which were selected because they are measures of order (entropy) or changes of pattern (LZC) of the ECG signals. This SVC machine is capable to get several clusters according to the kind ECG signals involved.

2. Support Vector Clustering

Minimal radio sphere: For a set of data points x_i, $i = 1, 2, \ldots, n$, from data space, and using a nonlinear mapping Φ (kernel function) from the data space to high dimensional feature space. Any function $\Phi(\cdot)$ that satisfies the Mercer's condi-

tions [15] is candidate to be a kernel function. The smallest sphere algorithm will find the minimal radius R subject to the constraints (1)

$$\|\Phi(x_i) - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0, \tag{1}$$

where $\|\cdot\|$ is the Euclidean norm, a is the center of the sphere, and ξ_i is a slack variable in the feature space to relax the solution which represents outliers in the data space.

The Lagrangian function of (1) is the equqtion

$$L = R^2 - \sum_{i=1}^{n} \left(R^2 + \xi_i - \|\Phi(x_i) - a\|^2\right)\beta_i - \sum_{i=1}^{n} \xi_i\mu_i + C\sum_{i=1}^{n} \xi_i \tag{2}$$

where $\beta_i \geq 0$ and $\mu_i \geq 0$ are Lagrange multipliers, C is a constant to take into account outliers, then $C\sum \xi$i is a penalty term.

Taking the derivative of (2) with respect to R, a and ξ_i and equaling to zero leads

$$\frac{dL}{dR} = 2R\left(1 - \sum \beta_i\right) = 0, \Rightarrow \sum \beta_i = 0, \tag{3}$$

$$\frac{dL}{da} = -\sum 2\Phi(x_i)\beta_i + 2a\sum \beta_i = 0, \Rightarrow a = \sum \Phi(x_i)\beta_i, \tag{4}$$

$$\frac{dL}{d\xi} = -\sum \beta_i - \sum \mu_i + \sum C = 0, \Rightarrow \beta_i = C - \mu_i. \tag{5}$$

The Karush-Kuhn-Tucker conditions [15] establish that:

$$\xi_i\mu_i = 0 \tag{6}$$

$$\left(R^2 + \xi_i - \|\Phi(x_i) - a\|^2\right)\beta_i = 0. \tag{7}$$

If a data point has $\xi_i > 0$ and $\beta_i > 0$, (6) implies that $\mu_i = 0$, then $\beta_i = C$. This data point will be called bound support vector (BSV), because its image is outside of the minimal sphere. If a data point has $\xi_i = 0$, its image lies inside of the sphere in the feature space, it can observe from (7) the term between parenthesis is not equal to zero. And if a data point has $0 < \beta_i < C$, then its image lies on the surface of the sphere, in this case the data point are called the support vectors (SV).

Dual form of (2) is obtained substituting Equations (4)–(6) in it

$$W = \sum_i \Phi(x_i)^2\beta_i - \sum_{i,j} \beta_i\beta_j\Phi(x_i)\Phi(x_i) \tag{8}$$

Because of variable μ_i does not appear in Equation (8), constraint (5) is changed to

$$0 < \beta_i < C, \tag{9}$$

constraint (3) keeps unchanged ($\sum \beta_i = 1$).

Kernel matrix is equal to inner product of two vectors $\Phi(x_i)$ and $\Phi(x_j)$ in the feature space, this is: $K(x_i, x_j) = \Phi(x_i) \cdot \phi(x_j)$. The advantage of using kernel matrix, the computation is done in an arbitrary feature space without explicitly using $\Phi(x)$. Kernel matrix using a Gaussian kernel function, as proposed by [16], is

$$K(x_i, x_j) = e^{\frac{\|x_i - x_j\|^2}{2\sigma^2}}, \tag{10}$$

where parameter σ is the width of the Gaussian.

Dual form (8) after substituting (10) is

$$W = \sum_i K(x_i, x_j)\beta_i - \sum_{i,j} \beta_i\beta_j \sum_i K(x_i, x_j). \tag{11}$$

The distance $R(\cdot)$ of any point at x to the center a in the feature space is [13]:

$$R^2(x) = \|\Phi(x_i) - a\| = \Phi(x_i) \cdot \Phi(x_i) - 2\Phi(x_i) \cdot a + a \cdot a \tag{12}$$

Taking into account (4), equation (12) left as

$$R^2(x) = K(x, x) - \sum_i \beta_i K(x_i, x) + \sum_{i,j} \beta_i\beta_j K(x_i, x_j). \tag{13}$$

Optimization problem consists on finding β values that maximizes (11) subjects to the constraints $0 < \beta_i < C$, and $\sum \beta_i = 1$. There are three possibilities in which each data point can lie with respect to the minimal sphere: $R^2(x) = R$ if x is a SV, $R^2(x) > R$ if x is a BSV, and $R^2(x) < R$ if x is inside of the minimal sphere. Figure 1 shows a representation of different data points in the minimal sphere.

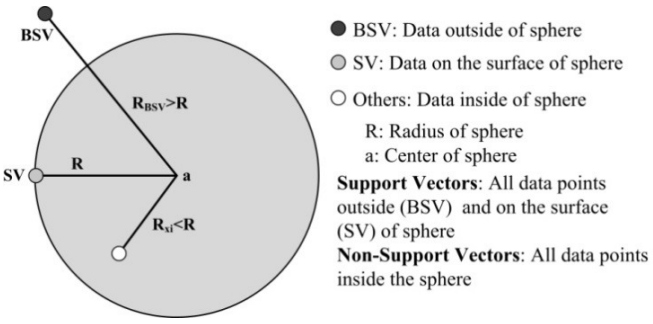

Figura 1: Representation of different data points in the minimal sphere.

One–Class classifier [17, 18]: The main objective of this learning machine is to find the farthest hyperplane from the origin in the feature space that separates all the data in the side (of hyperplane) opposite to the origin.

$$min\frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\Phi_i - \rho, \tag{14}$$

subject to constraints $w^T\Phi(x_i) \geq \rho - \xi_i, \xi_i \geq 0$ and, $\rho \geq 0$; where w is the norm of the hyperplane, ρ is the distance of the hyperplane to the origin, and ν controls the influence of outliers. The Lagrangian function of (14)

$$L = \frac{1}{2}w^T w + \frac{1}{\nu n}\sum_{i=1}^{n}\xi - \rho - \sum_{i=1}^{n}\alpha_i\left(w^T\Phi(x_i) + \xi_i - \rho\right) - \sum\beta_i\xi_i, \tag{15}$$

subject to $\alpha_i, \beta_i \geq 0$.

Taking the derivative of (15) with respect to w, ξ_i, ρ and equaling to zero leads

$$\frac{dL}{dw} = 0 \quad \Rightarrow \quad w = \sum\alpha_i\Phi(x_i), \tag{16}$$

$$\frac{dL}{d\xi_i} = 0 \quad \Rightarrow \quad \alpha_i + \beta_i = \frac{1}{\nu n} =, \tag{17}$$

$$\frac{dL}{d\rho} = 0 \quad \Rightarrow \quad \sum\alpha_i = 1. \tag{18}$$

Substituting equations (16)–(18) in (14) and taking into account $K(x_i, x_j) = \Phi(x_i)\cdot\Phi(x_j)$ leads to the dual form of (14)

$$min\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j\Phi^T(x_i)\cdot\Phi(x_j) = min\frac{1}{2}\sum_{i}\alpha_j\alpha_j K\left(x_i x_j\right), \tag{19}$$

The Karush-Kuhn-Tucker conditions [15] establish that:

$$\alpha_i\left(w^T\Phi(x_i) + \xi_i - \rho\right) = 0, \tag{20}$$

$$\beta_i\xi_i = \left(\frac{1}{\nu n} - \alpha_i\right)\xi_i = 0. \tag{21}$$

Table 1: Summary of conditions defining Support Vectors and non-SV.

Support Vector	Condition	Distance
BSV	$\alpha_i = \frac{1}{\nu n}, \xi_i > 0$	$w^T\Phi(x) < \rho$
Non-Bound Support Vector	$0 < \alpha_i < \frac{1}{\nu n}, \xi_i = 0$	$w^T\Phi(x) = \rho$
Non-Support Vector	$\alpha_i = 0, \xi_i = 0$	$w^T\Phi(x) = \rho$

Table 1 summarizes several conditions that define when a data point is a SV and when is a non-SV, considering equation (20) and (21).

Cluster Assignment: Cluster algorithm is not capable to know if several points belong to different clusters [14]. Then is necessary to do a cluster assignment to each point, this assignment process consists on using a geometric approach involving the radius of minimal sphere in the feature space [14]. It is well known, that the cluster assignment is an expensive process, then Lee and Daniels have proposed an efficient cluster labeling called cone cluster labeling [13] in which they define a cone between two support vectors (support vector cone), if two or more cones are intercepted the support vectors that define each cone belongs to the same cluster, that is, the maximum number of cluster corresponds to the number of SVs. Each cone is defined by the angle that forms two support vectors. In [13] authors demonstrated that angle between two

vector images Φ(xi) and Φ(xj), when the kernel function is a Gaussian, is defined by

$$K\left(x_i, x_j\right) = \Phi(x_i) \cdot \Phi(x_j) = \cos(\theta), \tag{22}$$

where θ is the angle between the vectors $\Phi(x_i)$ and $\Phi(x_j)$. Equation (10) demonstrates that distances in data space corresponds to angles in feature space and vice versa. This last affirmation is important because the clustering process is done in the data space using the distance that corresponds of its image (angle) in the feature space. Lee and Daniels have demonstrated [13] that when the function kernel is a Gaussian the support vectors which lie on the minimal sphere correspond to the support vector that lie on the optimal plane in the feature space for the one-class classifier, that is, support vectors lie on the interception of the optimal hyperplane with the minimal sphere. For this reason the machine selected to find the support vectors is the one-class classifier.

3. Lempel-Ziv Complexity

LZC is a measure of the number of different patterns in a finite sequence. The process to determine the complexity of a signal consists on transforming it into a finite sequence made up of just a few symbols [8, 19]. Discrete-time biomedical signal is converted into a binary sequence, by comparing it with a threshold (Td) [20]. Given a discrete signal $X = x(1), x(2), \ldots, x(n)$, it is converted into a finite sequence $S = s(1), s(2), \ldots, s(n)$ of 1's and 0's as

$$s(i) = \begin{cases} 0, & \text{if } x(i) < T_d \\ 1, & \text{if } x(i) \geq T_d \end{cases}, \quad i = 1, 2, \ldots, n. \tag{23}$$

To compute LZ complexity of a sequence, this is scanned from left to right and the complexity counter $c(n)$ is increased by one unit when a new pattern of consecutive characters is found [19]. For example, Lempel-Ziv complexity of s = 001111000011100001111001100011110 is 7, because different patterns observed in s are 0|01|1110|0001|1100001111|00110|0011110.

An upper bound of $c(n)$ [19, 20] is

$$b(n) = \frac{n}{\log_m(n)}, \tag{24}$$

where m is the number of different symbols given in (22).

The normalized LZC $C(n)$ of a arbitrary random sequence of length n is [20, 21]

$$C(n) = \frac{c(n)}{b(n)} \tag{25}$$

4. Shannon's Entropy

Entropy is a measure that quantifies the level of randomness of a signal. As an intuitive idea, the entropy of a system is proportional with the logarithm of the number of its possible states. Shannon established the SE as the way to measure the amount of uncertainty by a probability distribution function on a finite set [10]. Shannon entropy of the sampled signal $X = x1, x2, \ldots, xn$ is written $H(X)$ and is defined by [11]

$$H(X) = -\sum^{n} p(x_i) \log p(x_i), \tag{26}$$

where p(·) is the probability distribution function, that assigns a probability between 0 and 1 to x_i.

5. Experimental Part

Preprocessing of ECG Signals: ECG data was obtained from the MIT-BIH Arrhythmia Database [22]. Each signal of the MIT-BIH Arrhythmia database was sampled at a frequency of 360 Hz. One normal ECG (record 100) and three ECG signals with arrhythmic beats (records 101, 105 and 109) were selected from the MIT-BIH Arrhythmia Database, see Figure 2. All those signals were preprocessed to remove muscle and breathing noise, electrode motion artifacts, power line interference and DC offset. ECG signals were divided in 140 segments of 35 four-second segments each of each ECG record. Training data was obtained from those segments of preprocessed ECG signals. Sample size (1428 samples) of the segment was selected to pick at least four cardiac cycles. The SE and LZC pair value by

each segment was obtained; Figure 3 shows the ECG training data in the space SE-LZC space, each point in Figure 3 represents a segment of ECG signal, and each segment is identified with a different symbol.

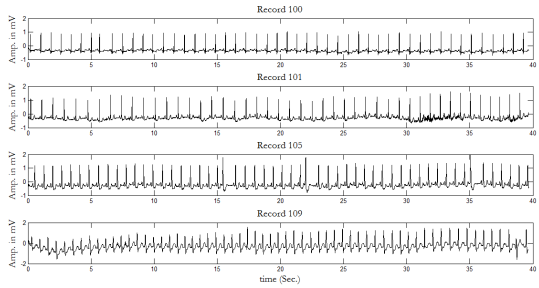

Figura 2: Forty-second segments of the MIT-BIH Arrhythmia Records.

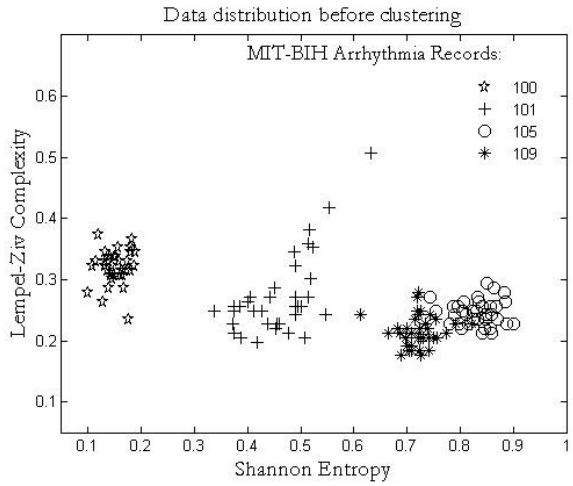

Figura 3: Data distribution before clustering.

Clustering of ECG Signals: SVC was trained using LIBSVM [23] toolbox. LIBSVM is based on an efficient algorithm that permits work with huge quantity of training data. Table 2 shows a summary of the several trials done to tune the SVC machine. As shown in Table 2, the number of clusters can be adjusted varying the parameters ν and σ of SVC and the maximal distance to determine a cluster. Third column of Table 2 is the distance of the optimal hyperplane to the origin in the feature space. Fourth column of Table 2 is the clustering distance, modified by a user-defined scale factor (radius modifier factor) for clustering in the SE-LZC plane, which corresponds to (a non-linear mapping of) the angle that defines the cone between two support vectors in the feature space [13], and last column shows the number of clusters obtained with those parameters.

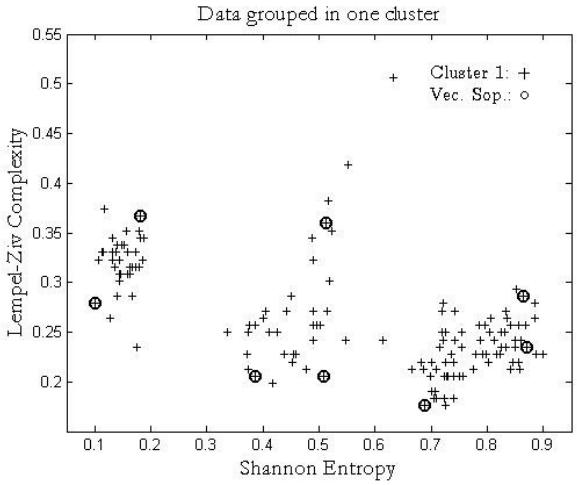

Figura 4: Clustering of ECG signals using $\nu = 0.001$, clustering distance = 0.2785, $\sigma = 0.2$.

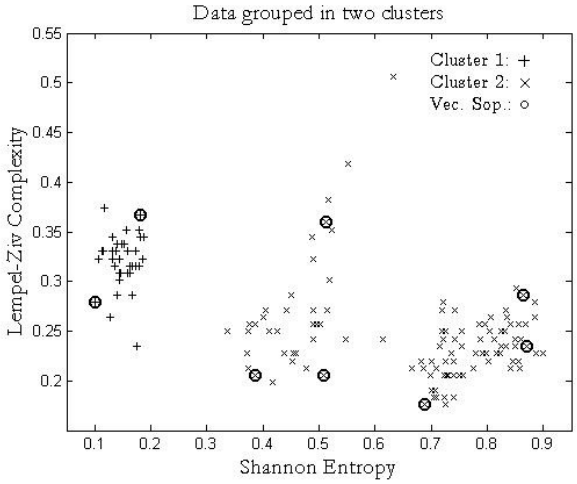

Figura 5: Clustering of ECG signals using $\nu = 0.001$, clustering distance = 0.1949, $\sigma = 0.2$.

Figures 4– 7 show the data point assignment to a cluster, identification symbols in Figures 4– 7 does not correspond to the symbols identifiying each ECG signal in Figure 3. Support vectors are those data points in a black circles.

Figure 3 is the patron distribution pattern to tune the SVC parameters (varying σ and ν, and

Table 2: Summary of parameter tuning of the Support Vector Clustering.

Nu ν	Sigma σ	Rho ρ	#Vs	Clustering distance	#Clusters
0.001	0.2	0.0303	8	0.2785	1
0.001	0.2	0.0303	8	0.2785	2
0.001	0.2	0.0303	8	0.2785	3
0.001	0.2	0.0303	8	0.2785	4

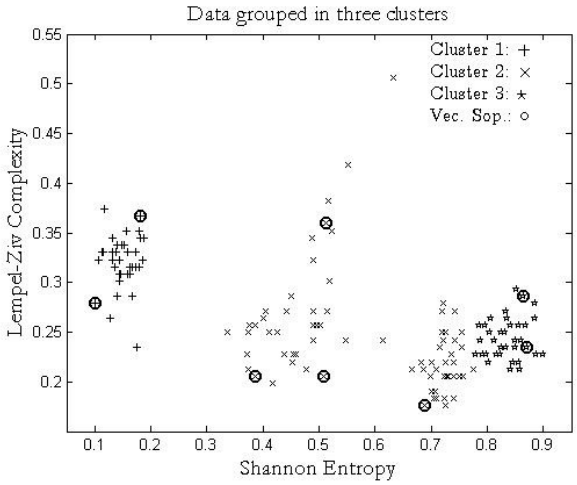

Figura 6: Clustering of ECG signals using $\nu = 0.001$, clustering distance = 0.1894, $\sigma = 0.2$.

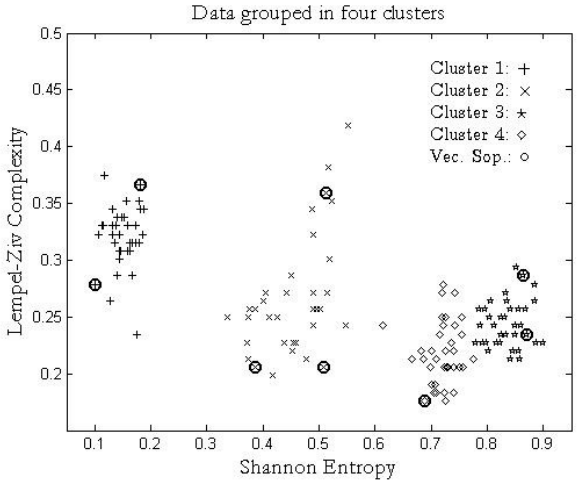

Figura 7: Clustering of ECG signals using $\nu = 0.001$, clustering distance = 0.1559, $\sigma = 0.2$.

the clustering radius) to get the suited number of clusters. Figure 4 shows as a first aproximation only one cluster which does not correspond, of course, to the real groups of the original data. This solution is the farthest approximation solution found to the real one.

Figure 5 shows a better approximation of the clusters. ECG record 100 was correctly identified as cluster 1 (+). ECG records 101, 105 and 109 were grouped in a cluster 2 (×), which can be understood as a pathological patterns.

Figure 6 shows clearly how the SVC succesfully separated ECG records 100 (cluster 1) and 101 (cluster 2) in two different clusters. ECG record 105 was assigned to cluster 2 and ECG record 109 was grouped in the cluster 3 identified with a star symbol.

Figure 7 shows the best solution found according to the original data distribution, see Figure 3. All the records were successfully grouped in different clusters. ECG record 105 was assigned to the cluster 4, which is identified with a diamond symbol. ECG record 109 was grouped in cluster 3, it was identified with a star symbol. Some points (ECG segments) of ECG record 109 were assigned cluster 4, because of the very close distance between these two groups. As you can note, the SVC machine for Figure 4 through Figure 7 is the same one, the unique difference is the clustering distance (non-linear mapping of the cone's angle in the feature space to the SE-LZC plane modified by a user-defined scale factor) used to assign the points to a cluster in the SE-LZC plane. The first step, before assign the entire non-support vector points to a group, is to cluster the support vectors

taking into account the new clustering distance.

It is important to highlight during the clustering process the following: a) The number of clusters was obtained manipulating the SVC parameters (σ and ν) and the clustering distance, b) It is important to count with asserted criteria of an expert to interpret the quality of the clusters obtained.

6. Conclusions

Results showed that SVC machine was capable to cluster successfully several normal and pathological ECG signals. Number of clusters was set by mean of several trials manipulating the SVC parameters (σ and ν) and the clustering distance, which is a scaled non-linear representation of the cone's angle in feature space. The cluster labeling of the rest of (non-support vector) data was done after the support vectors were clustered, that is, process of cluster labeling is less expensive compared with another clustering methods, because the cluster assignment of each point is done after all cluster are defined, and that assignment depend on the distance of one point to a support vector defining a cluster. The overall accuracy of the data point assigning was 95.7 %.

Referencias

[1] Claudia Lerma, Trine Krogh-Madsen, Michael Guevara, and Leon Glass. Stochastic aspects of cardiac arrhythmias. *Journal of Statistical Physics*, 128(1):347–374, 2007.

[2] Gari D Clifford, Francisco Azuaje, Patrick McSharry, et al. *Advanced methods and tools for ECG data analysis*. Artech house London, 2006.

[3] Roger G Mark and George B Moody. Ecg arrythmia analysis: Design and evaluation strategies. In I. Gath and G. F. Inbar, editors, *Advances in Processing and Pattern Analysis of Biological Signals*, chapter 18, pages 251–272. Springer, 1996.

[4] Douglas A Coast, Richard M Stern, Gerald G Cano, and Stanley A Briller. An approach to cardiac arrhythmia analysis using hidden markov models. *Biomedical Engineering, IEEE Transactions on*, 37(9):826–836, 1990.

[5] Sophia H Zhou, Pentti M Rautaharju, and Harry P Calhoun. Selection of a reduced set of parameters for classification of ventricular conduction defects by cluster analysis. In *Computers in Cardiology 1993, Proceedings.*, volume 20.

[6] S Barro, R Ruiz, D Cabello, and J Mira. Algorithmic sequential decision-making in the frequency domain for life threatening ventricular arrhythmias and imitative artefacts: a diagnostic system. *Journal of biomedical engineering*, 11(4):320–328, 1989.

[7] Andrei N Kolmogorov. Three approaches to the quantitative definition of information. *Problems of information transmission*, 1(1):1–7, 1965.

[8] Abraham Lempel and Jacob Ziv. On the complexity of finite sequences. *Information Theory, IEEE Transactions on*, 22(1):75–81, 1976.

[9] Claude E Shannon. Communication theory of secrecy systems. *Bell system technical journal*, 28(4):656–715, 1949.

[10] George J Klir. *Uncertainty and information: foundations of generalized information theory*. John Wiley & Sons, 2005.

[11] Jianbo Gao, Jing Hu, and Wen-wen Tung. Entropy measures for biological signal analyses. *Nonlinear Dynamics*, 68(3):431–444, 2012.

[12] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[13] Sei-Hyung Lee and Karen M Daniels. Gaussian kernel width selection and fast cluster labeling for support vector clustering. Unpublished sc. d. dissertation, University of Massachusetts Lowell, Department of Computer Science, 2005.

[14] Asa Ben-Hur, David Horn, Hava T Siegelmann, and Vladimir Vapnik. Support vector clustering. *The Journal of Machine Learning Research*, 2:125–137, 2002.

[15] Vladimir Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2000.

[16] David MJ Tax and Robert PW Duin. Support vector domain description. *Pattern recognition letters*, 20(11):1191–1199, 1999.

[17] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

[18] Bernhard Schölkopf, Robert C Williamson, Alex J Smola, John Shawe-Taylor, and John C Platt. 18. In Sara A. Solla, Todd K. Leen, and Klaus-Robert Muller, editors, *Advances in Neural Information Processing Systems, NIPS*, volume 12, pages 582–588, 1999.

[19] Lisheng Xu, David Zhang, Kuanquan Wang, and Lu Wang. Arrhythmic pulses detection using lempel-ziv complexity analysis. *EURASIP Journal on Advances in Signal Processing*, 2006(1):018268, 2006.

[20] Mateo Aboy, Roberto Hornero, Daniel Abásolo, and

Daniel Álvarez. Interpretation of the lempel-ziv complexity measure in the context of biomedical signal analysis. *Biomedical Engineering, IEEE Transactions on*, 53(11):2282–2288, 2006.

[21] Ana Carolina Sousa Silva, Sérgio Souto, Euvaldo Ferreira Cabral, and Ernane José Xavier Costa. Using an easy calculable complexity measure to introduce complexity in the artificial neuron model. *Research Journal of Biological Sciences*, 2(5):607–611, 2007.

[22] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[23] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.