SHAPE-BASED CLUSTERING OF ELECTROCARDIOGRAM TIME SERIES WITH DYNAMIC TIME WARPING

Alessandro Montemurro

Technical University of Denmark, DTU Compute

***Abstract* – This paper presents a clustering algorithm for electrocardiogram time series. The purpose is to understand if it is possible to detect the gender of an individual using the only electrocardiogram time series. The classification will be carried out without giving any additional information about the individuals. Hence, the algorithm is based only on the shape of the electrocardiograms. A suitable similarity measure, capable of explaining the similarity between time series, has to be defined. This similarity measure has to be flexible because of the inter-subject variability. For this purpose, Dynamic Time Warping is used to sort individuals according to Male/Female. It will be shown that, to some extent, it is possible to automatically characterize a male (or female) heartbeat based on its shape.**

***Index Terms* —** ECG, machine learning, hierarchical clustering, DTW, gender detection

1. INTRODUCTION

Healthcare researchers and medical institutions turned to machine learning to handle the significant amount of available data and automate the diagnostic process. Electrocardiography (ECG) is one of the fields where machine learning can be a powerful tool. An electrocardiogram is a record of the magnitude of the electrical forces inside the heart. ECGs can be seen as time-series where the observations are both spatially and time correlated. Closely related to ECG, the vectorcardiogram (VCG) is a 3D representation of the depolarization (depolarization cycle) of the heart; it calculates the magnitude and direction of the electrical signals emanated from the heart. These vectors are used to make three projections of the polarization event of the heart, namely coronal (frontal), transverse (horizontal) and sagittal (vertical) plane. VCG uses a continuous series of vectors that form curving lines around a central point to describe the electrical activity of the heart. [1]

The purpose of the study is to apply data mining to ECGs and VCGs and try to discover underneath patterns in the data. An unsupervised clustering algorithm is used to group similar VCGs. In particular, it is interesting to understand to which extent it is possible to assess the gender on an individual, based only on the shape of the waves in the ECG, without giving any other information about the patient.

Electrical activity causes contraction of the cardiac muscle. As in neurons or muscles, an electrical impulse is generated in the heart by the depolarization of the cell membranes. This phenomenon is due to the passage of ions Na^+ inside the cell and ions K^+ outside. The impulse is generated by the auto-rhythmic cells of the sinoatrial node, the natural pacemaker of the heart, that is placed in the right atrium. It spreads towards the atrioventricular node; this process depolarizes the atria, causing their contraction. The blood is pumped in the ventricles. The impulse then propagates into the ventricles, causing their polarization (contraction). The blood is finally pumped outside the heart. A potential difference causes all the different phases of the cardiac cycle. The electrocardiogram shows the sum of potentials generated by all the cardiac cells. [2]

In order to measure the potential difference, some electrodes are used to detect the potential on the surface. The different cardiac phases correspond to different waves in the ECG, as shown in Figure 1. Depolarization moving towards a lead direction causes a positive, or upwards, deflection on the ECG. Depolarization moving away from a lead direction causes a negative, or downwards, deflection.

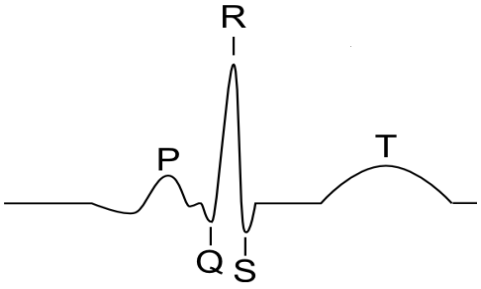

Figure 1. Representation of the different waves in the ECG. From left to right: P-wave (atria polarization), QRS complex (ventricles polarization), T-wave (ventricles depolarization). The depolarization of the atria in hidden by the big QRS complex.

Vectorcardiography is presently used mostly for didactic purposes to teach physiological aspects of electrocardiography. The VCG and the 12-lead ECG

represent the same information, albeit in different formats. The principal advantage of the VCG is that it provides the same information as the 12-lead ECG but with fewer leads. This is achieved, as mentioned, by manipulating these orthogonal vector signals to yield a conventional ECG signal. VCG is not used in clinics because of the expensiveness of the equipment and the problematic interpretability. However, it is believed that the analysis of VCGs can highlight some features in the heartbeat which could not be detected using other ECG parameters. [4]

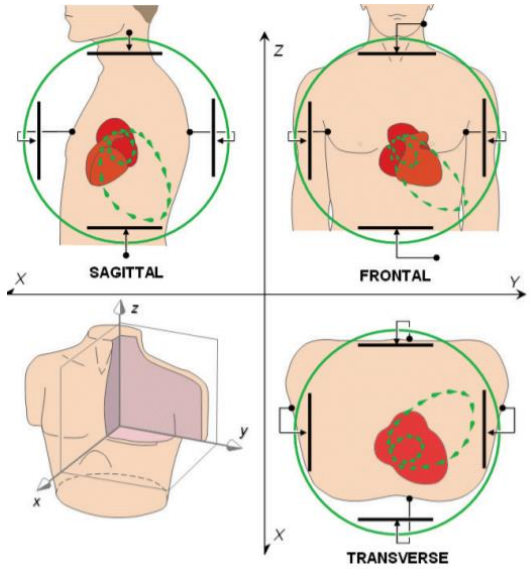

Figure 2. The view of VCG in three Cartesian planes. The dashed green line represents the VCG. [3]

The objectives of this study are:

- Define a distance measure able to capture the similarity between ECG data;
- Extend the concept of clustering to time series;
- Group similar ECGs according to their shape using an unsupervised algorithm.

2. DATA

The data analyzed in this study were provided by Glostrup Hospital, Copenhagen. The dataset consists of N = 6667 patients. For each patient, the following information are available:

- 1.2 seconds recording of a single beat of a 12-lead ECG. This is obtained by averaging a longer 10 seconds recording. Sampling frequency = 500 Hz;
- General information: Age, sex, duration of the follow-up, whether the patient survived or not, whether the patient got diagnosis of Ischaemic Heart Disease (IHD);
- Clinical data: Heartbeat, QRS duration, T-wave peak, Body Mass Index (BMI).

Together with the recordings, the ECG machine also provides the starting (onset) and the ending (offset) samples of each phase of the heartbeat. Thanks to these markers, it is possible to extract single waves and focus on them separately.

New information can be extracted from the available data: three dimensional VCG can be obtained by 12-leads ECG passing the signal through the *Dower transformation* [5]. Fig. 2 shows a view of the VCG signal from three Cartesian planes (XY, YZ and XZ). Fig. 3 shows the vector loops for P, QRS, and T-wave activities**.** The largest green QRS-loop manifests the ventricular depolarization activities. Red P wave is the atrial depolarization after the SA node excitation. The ventricular repolarization is shown as blue T-loop [6].

Another useful quantity to describe the heart activity is the magnitude lead. One dimensional magnitude lead of VCGs can be computed as the Euclidean norm of the VCG:

$$magnitude\_lead = \sqrt{VCG_x^2 + VCG_y^2 + VCG_z^2}$$

where VCG_i, $i = \{x, y, z\}$, are the three spatial components of the VCG. One dimensional magnitude lead is useful when we need to lower the dimension of data, transforming a three dimensional time series into one dimensional. Since magnitude lead is defined as a norm, it gives information about the magnitude of the beat; however it makes us lose the information about the direction: for each time point, the magnitude of the VCG will be a scalar and not a spatial vector.

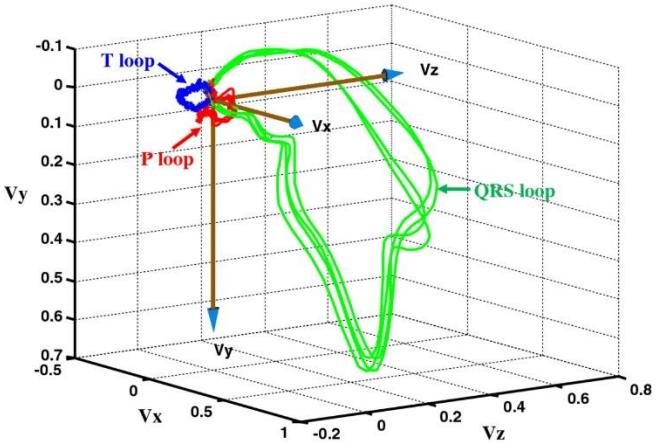

Figure 3. A representative VCG plot [6].

3. METHODS AND ALGORITHM

The basic idea of the algorithm proposed in this paper is to group patients in different clusters using the only median

ECG recording. All the other information about the patients are used later, to examine each cluster.
Since clustering is a distance-based algorithm, the problem one faces is to understand the meaning of similarity between ECG. Mathematically, this problem turns into the one of choosing a suitable similarity measure capable of explaining the similarity between time series. Common distance measures, e.g. Euclidean or L_p distances, are not suitable since they require that the two time series have the same length, in terms of number of samples. Moreover, they are very sensitive to signal transformations such as shifting, uniform amplitude scaling and uniform time scaling. Due to inter-subject variability, the duration and amplitude of the phases of the heart beat can be different; hence, the measure must be chosen so that time series with different length are handled.

An algorithm called *Dynamic Time Warping* (DTW) is used to compute the similarity between ECGs and, on top of it, a hierarchical clustering algorithm is built.

3.1. Dynamic Time Warping (DTW)

Dynamic Time Warping (DTW) is a time series alignment algorithm originally developed for speech recognition by D. Berndt and J. Clifford [7]. In general, DTW is an algorithm for measuring the similarity between two time series which may vary in speed. It calculates an optimal match between two time series. The sequences are warped non-linearly in the time dimension to determine a measure of their similarity independent of specific non-linear variations in the time dimension. In addition to the similarity between the two sequences, a *warping path* is returned: by warping the two series according to this path, they may be aligned in time. The main feature of this distance measure is that it allows recognizing similar shapes, even if they present signal transformations, such as shifting or scaling. Furthermore, time series of different length can be compared, because DTW replaces the one-to-one point comparison, used in the Euclidean distance, with a many-to-one and one-to-many comparison. Therefore, even if the two sequences have different length, one point of the first sequence can be associated with many points of the other.

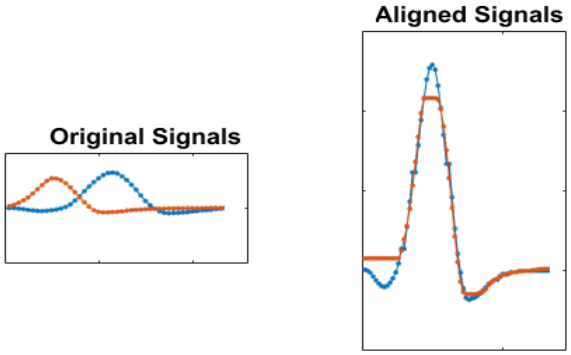

Figure 4: (left) original QRS signals; (right) aligned version of the signals using DTW optimal match.

Another essential aspect of DTW algorithm is that it can handle multidimensional time series and compute the similarity between them. In this study, this is a fundamental requirement since the electrical activity of the heart is described by ECG, which is a 12 dimensional time series (each lead represents a dimension) and VCG, which is three-dimensional.

The pseudo-algorithm for the construction of the optimal path is presented below.

Algorithm 1

Input S: time series with length n;
T: time series with length m.

Output *DTW* matrix

1. Initialize the $n \times m$ matrix *DTW*;
2. for $i = 0 : n$
3. for $j = 0 : m$
4. $DTW[i, j] = |S[i] - T[j]| + min(DTW[i-1, j], DTW[i, j-1], DTW[i-1, j-1])$
5. end
6. end

The *DTW* matrix is used to find the optimal path between the sequences *S* and *T*. The actual distance *dtw* (*T*, *S*) is given by the element *DTW[n, m]*. This is due to the fact that, in each loop iteration, the distance between the aligned sequences is accumulated. This is expressed in step 4 of Algorithm 2. At entry $[i, j]$ of the matrix, also the contribution of previous iterations is involved in the computation.
This measure between time series will be used in the following step, to perform clustering.

Fig. 4 shows an example of DTW optimal match. The two original sequences have a similar shape, but they are "out-of-phase." If a point-to-point Euclidean distance is used to assess the similarity of the two time series, they will be classified as very different sequences because one is time-shifted respect to the other. Indeed, because of this shifting, we would like to assess that they are similar. Right panel plot of Fig. 4 shows the time aligned signal using the DTW optimal match. Now the series are aligned and we can appreciate the similarity in their shape. After having aligned the sequences, it is possible to compute their distance: only the "real distance" between them will be taken into account, and the distance due to the time shifting will be ignored.

3.2. Hierarchical clustering

Once a similarity measure is identified, similar observations can be clustered together. *k*-means clustering is not suitable for this problem because it requires to specify the centers of

the k clusters and it's not trivial to define a *mean ECG* for each cluster. Hence, a hierarchical agglomerative clustering (HAC) fits this problem since it does not requires specifying the number of clusters and the mean vectors of the clusters. Instead of finding one single k, HAC creates a nested sequence of partitions organized as a hierarchy. Agglomerative hierarchical clustering is said to be a bottom-up approach: the bottom of the hierarchy corresponds to the finest partition (each observation is a singleton) whereas the top-level of the hierarchy corresponds to the coarsest possible partition corresponding to putting every observation in the same cluster. [8]
A hierarchical clustering algorithm proceeds with the following steps :

Algorithm 2

1. Initialize the $N \times N$ distance matrix D;
2. For all $\forall\, i, j$, compute $[D]_{i,j} = dtw(ECG_i, ECG_j)$;
3. Place each observation in a separate cluster;
4. Iteratively, *merge* the two closest clusters;
5. Repeat until all observations are in the same cluster.

In step 2, the distance matrix is filled with the distances of all the possible pairs; the distance function depends on the particular application. In this case, the DTW measure is used (see Section 4.1):

$$[D]_{i,j} = dtw(ECG_i, ECG_j) \qquad i, j = 1 \dots N$$

where ECG_i is the ECG recording for patient i. By definition of distance, D is a symmetric matrix since

$$dtw(ECG_i, ECG_j) = dtw(ECG_j, ECG_i) \qquad \forall\, i, j.$$

In addition to the concept of distance between points, a key concept in hierarchical clustering is the distance *between cluster*. In step 4 of Algorithm 2, the two *closest* clusters are merged. The distance between clusters used in this algorithm is the *Ward's distance.*
Suppose at a given step of the algorithm there are K clusters. For each cluster, the centroid μ_k is computed as the average of the time series in cluster k; μ_k is a time series itself.
Define the error function

$$E = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{i,k} ||ECG_i - \mu_k||_2^2 \qquad (1)$$

$$z_{i,k} = \begin{cases} 1 & \text{if patient } i \text{ belonges to cluster } k \\ 0 & \text{otherwise} \end{cases}$$

Equation (1) represents the sum of squared errors between each time series and the clusters it is assigned to. According to Ward's method, the two clusters that provides the *smallest* increase in the above error are then merged. In other words, Ward's method tends to merge the clusters whose *merging cost* is minimum in terms of residual error. [8]
The algorithm described above, provided with DTW distance for time series and Ward's distance for clusters, produce the hierarchy of clusters. Such hierarchy can be synthesized in a tree-diagram, also called *dendrogram*.
Since HAC consists of a nested sequence of clusters, it is possible to choose the number of clusters after the algorithm is performed. Selected the desired number of clusters, the dendrogram can be cut to the corresponding level and the clusters are obtained.

4. EXPERIMENTS

The algorithm described in the previous paragraph has been applied to Glostrup dataset. The software used for the analysis is R v. 3.4.4, R Core Team, Vienna, Austria.

Define the QRS magnitude lead as the magnitude of the only QRS-loop. It is possible to extract the only QRS-loop from the VCG because the ECG machine provides the onset and offset of the QRS complex in the ECG; these points are the same in the VCG and give the starting and ending samples of the QRS-loop.
First, the algorithm is applied to the QRS magnitude lead: the data dimension is compressed from three to one. Then, the multidimensional version of DTW algorithm is used to cluster the three dimensional VCGs: QRS-loop and T-loop. The multidimensional DTW is tested in even higher dimension: individuals are clustered using the 12-leads T-wave of the ECG, where each lead represents a dimension.
For all the experiments mentioned before, a dendrogram is built and then cut to different levels; the chosen levels are 4, 6 and 10, i.e. the patients are grouped, respectively, in 4, 6, and 10 clusters.

To summarize, HAC has been performed on the following data:

1. QRS magnitude lead from the VCG (`vcg_mag_qrs`);
2. QRS-loop from VCG (`vcg_qrs`);
3. T-loop from VCG (`vcg_tloop`);
4. 8 leads T-wave from ECG (`ecg_twave`).

5. RESULTS

The output of the algorithm described in Section 4 is a hierarchy of clusters; once we specify the desired number k,

Cluster n.	1	2	3	4	p-value
vcg_mag_qrs	56.3	43.8	41.2	73.0	<0.001
vcg_qrs	40.6	52.0	44.8	61.7	<0.001
vcg_tloop	**18.7**	51.4	**85.6**	63.2	<0.001
ecg_twave	43.4	76.6	**19.0**	51.7	<0.001

Cluster n.	1	2	3	4	5	6	p-value
vcg_mag_qrs	58.9	53.5	43.8	42.9	38.6	73.0	<0.001
vcg_qrs	66.0	**17.7**	43.7	64.4	44.8	61.7	<0.001
vcg_tloop	**19.7**	43.5	**17.0**	**85.6**	63.2	70.4	<0.001
ecg_twave	38.0	51.1	**82.6**	**19.0**	70.8	51.7	<0.001

Cluster n.	1	2	3	4	5	6	7	8	9	10	p-value
vcg_mag_qrs	54.7	63.0	53.5	40.0	42.9	47.8	38.9	37.7	72.0	**80.5**	<0.001
vcg_qrs	45.0	73.3	**17.7**	48.0	64.4	37.7	23.2	60.1	63.0	65.8	<0.001
vcg_tloop	21.7	46.8	**17.0**	38.1	**16.6**	**86.3**	**84.7**	72.0	55.0	70.4	<0.001
ecg_twave	32.7	71.9	27.3	**82.6**	61.9	**15.9**	58.0	**20.7**	70.8	51.7	<0.001

Figure 5. The three tables summarize the results of all the experiments. The most interesting findings, discussed later in section 7, are highlighted with boldface.

we get *k* clusters. The next step is to analyze each cluster and try to extract some features that characterize it. In particular, since the aim of this study was to understand to which extent it is possible to characterize the shape of the ECG waves in terms of gender, we analyze the sex of the individuals *within the same clusters*.

The results are presented in Table 1. For each experiment, the percentage of males is reported. For example, in the first table, the first row (`vcg_mag_qrs`) says that the patients were clustered in 4 groups according to the magnitude lead of the QRS-loop. In the fists cluster, 53.3% are males, in the second 43.8%, and so on.

To assess the significance of the test, also the p-value is given in the tables; it is always smaller than 0.001.

6. DISCUSSION

Using both the QRS-loop magnitude lead and the 3D QRS-loop, all the clusters present an equal percentage of males and females, around 50%. This means that the algorithm is not capable of detecting gender and place patients with different sex in different clusters using the information supplied by QRS-loop in VCG.

The results are different when T-loop from VCG and T-wave form ECG are used: the obtained clusters are more *pure* in terms of separation of gender. For instance, if we use `vcg_tloop` and divide the observation into four clusters (see Table 1, first row of the upper table), cluster 1 has 18.7% of males; instead cluster 3 has only 14.4% (100-85.6) of females. This behavior was founded in all the other experiments, using T-loop and T-wave. This means that, using the only information provided by the depolarization of the ventricles (T-wave), it is somehow possible to differentiate the gender of the individual.

It is important to emphasize that the algorithm uses the only ECG or VCG to differentiate gender, with no additional information. This means that it "detects" gender based only on how, qualitatively, the ventricles relax. Hence, the algorithm is given only a sequence of numbers and it is able to stratify data (i.e. divide population into homogeneous groups) according to the qualitative contraction of the cardiac muscle.

It is also worth mentioning that dynamic time warping relies on some assumptions more or less reasonable. The main assumption is that one time series is a non-linear time-stretched version of the other. This means that the two time series are generated by the same process and, in one of them, the time-axis is non-linearly expanded and compressed. However, this assumption seems to be reasonable since two different ECGs will behave qualitatively in the same manner, unless some anomaly is present in the heart. Moreover, to overcome this strict limitation, DTW allows the two time series to have different numerical values.

7. CONCLUSION

In this paper, it has been shown that ECG and VCG can be useful to characterize the heartbeat of an individual based only on the shape of the waves. In particular, the focus is on the differentiation of ECGs and VCGs based on the gender of the patients. Unsupervised hierarchical clustering has been

used to stratify the data. To achieve this goal, Dynamic Time Warping distance measure has been used. This measure is able to explain the similarity of ECG data taking into account also the inter-subject variability, e.g., different duration of cardiac phases or different amplitude of them. The main results have been achieved with T-wave and T-loop, that correspond to the relaxation of the ventricles. In this case, patients are sharply separated in different clusters according to Male/Female.
Future works can include, for instance, the classification for *new* individuals using T-wave and T-loop as biomarkers.

8. REFERENCES

[1] G. E. Arrobo; C. A. Perumalla; Y. Liu; et al. "A Novel Vectorcardiogram System". University of South Florida.

[2] D. Dublin, "Interpretazione dell'ECG". Monduzzi Editoriale, 2018.

[3] J. Malmivuo and R. Plonsey, "Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields". Oxford University Press.

[4] G Daniel; G Lissa; D Medina Redondo; et al. (2007). "Real-time 3D vectorcardiography: An application for didactic use". Journal of Physics: Conference Series.

[5] M.S. Guillem; A.V. Sahakian; S. SwirynD (2006)"Derivation of Orthogonal Leads from the 12-Lead ECG. Accuracy of a Single Transform for the Derivation of Atrial and Ventricular Waves". Computers in Cardiology, no. 33, pp. 249-252.

[6] H. Yang et al. (2012). "Spatio-temporal representation of vectorcardiograms (VCG) signals". BioMed Central ltd.

[7] D. J. Berndt: J. Clifford (1994). "Using Dynamic Time Warping to Find Patterns in Time Series", IEEE.

[8] T. Herlau; M. N. Schmidt; M. Mørup (2017). "Introduction to Machine Learning and Data Mining". Technical University of Denmark.