Segment Clustering for Holter Recording Analysis

Jose Luis Rodríguez-Sotelo, Diego Peluffo-Ordóñez, David Cuesta-Frau, Germán Castellanos-Domínguez and Lina Sepúlveda*

Departamento de eléctrica, electrónica y computación, UNAL, Manizales, Colombia

Holter electrocardiography (ECG) is useful to detect transitory and irregular pathologies that are hard to diagnose in short-time ECG. The problem of this test is the wide amount of information which complicates the manual inspection. For this reason, computer analysis tools have been developed and are commonly used as diagnostic support. In general, these systems work off-line taking into account several factors that affect signal quality. Then, it is necessary to process each heartbeat, to detect some specific types of heartbeats. Therefore, unsupervised classification is preferred in this approach, being clustering the most frequently used technique. In addition, processing time and unbalanced classes are other important issues to be taken into consideration.

In this work, a methodology for segment grouping is presented that reduces the computational cost and sensitivity to unbalanced classes. Proposed method includes stages for preprocessing, characterization and clustering. All these stages are developed in a sequential scheme, where similar clusters are grouped into new clusters per couple of contiguous segments taking into account exclusion and merger criteria based on dissimilarities. For unsupervised grouping, the estimation of the number groups employing spectral techniques is first performed. Next, to avoid the problem of convergence into a local minimum, JH-means and sum of squares criteria are used. Clustering stage is developed by using a general iterative model for center-based clustering with soft membership functions.

The method is assessed over a set of records from MIT/BIH arrhythmia database with different types of heartbeats recommended by the AAMI, such as, normal, ventricular extra systoles, branch bundles blocks among others. The results are assessed by means the sensitivity and specificity measures, taking advantage of the database labels. Also, unsupervised performance measures are used. Finally, the performance of the algorithm is in average 95%, improving results reported by previous works of the literature.