

Clasificación de arritmias en registros Holter empleando análisis no supervisado de tipo particional y espectral

D. Peluffo^a, J. L. Rodríguez^a, and G. Castellanos^a

^aUniversidad Nacional de Colombia - Manizales

Resumen Se propone una metodología para el análisis no supervisado de patrones descriptivos de latidos en señales ECG de registros Holter, con el fin de clasificar arritmias y mostrar la aplicabilidad y versatilidad de las técnicas de análisis no supervisado. El sistema propuesto incluye etapas de caracterización, selección de características, criterios de inicialización, estimación del número de grupos y clustering, desarrolladas con técnicas de tipo particional y espectral. Los experimentos se realizan con la bases de de datos MIT/BIH.

Palabras clave: Análisis no supervisado, análisis espectral, arritmias, registros Holter

1. Introducción

Uno de los mayores problemas de la evaluación diagnóstica ambulatoria es la longitud de los registros Holter, lo que hace que la inspección visual sea un trabajo laborioso para los especialistas. Por esta razón, se han desarrollado herramientas de asistencia diagnóstica basadas en técnicas de procesamiento digital de señales y reconocimiento de patrones. En este caso, la idea general de la asistencia diagnóstica es disminuir la cantidad de latidos que debe revisar el especialista, de manera que, en lugar de revisar todos los latidos, únicamente deba revisar los latidos representativos de cada grupo obtenido por el sistema de análisis y, al final, sugerir un tipo de patología asociada. Para esta tarea se prefieren los métodos no supervisados porque no es necesario etiquetar el conjunto de datos y el proceso es más flexible, además, es la técnica de uso más frecuente en el diseño de dispositivos de análisis de registros Holter.

En este trabajo se presenta una metodología para el análisis no supervisado de patrones descriptivos de latidos en señales ECG de registros Holter, orientado a la clasificación de arritmias usando técnicas de análisis no supervisado. El sistema propuesto incluye etapas de caracterización, selección de características, criterios de inicialización, estimación del número de grupos y clustering. El conjunto de datos corresponde a características que han registrado buen desempeño en la literatura. Con el fin de mejorar la convergencia de la clasificación, se llevan a cabo etapas de estimación del número de grupos, selección de características e inicialización basadas en técnicas espectrales. Las pruebas se realizan sobre la base de datos de arritmias del MIT, considerando las recomendaciones de la AAMI. El desempeño del clustering fue evaluado en términos de sensibilidad, especificidad y selectividad. El método propuesto presenta resultados comparables con los reportados en la literatura.

2. Materiales y métodos

En el diagrama de bloques de la figura 1 se muestran las etapas de la metodología de análisis de ECG propuesta. La clasificación no supervisada incluye estimación del número de grupos e inicialización.

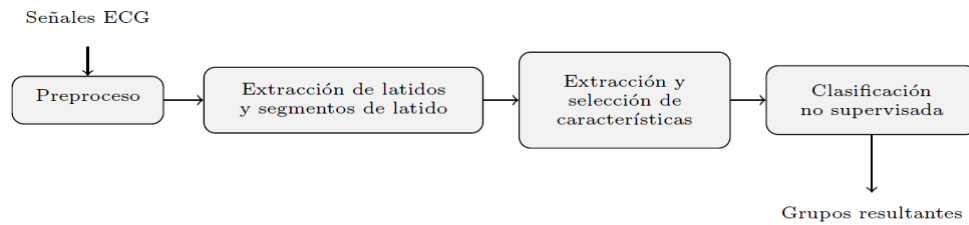


Figura 1. Diagrama de bloques de la metodología propuesta

2.1. Señales ECG

Las señales empleadas en este estudio fueron obtenidas de registros de la base de datos de arritmias del MIT (*Massachusetts Institute of Technology*), denominada MIT/BIH. Esta base de datos dispone de 48 registros con diferentes tipos de arritmia (A, L, R, V, entre otros). Las señales fueron adquiridas con una frecuencia de muestreo de 360 hz. Estos registros se encuentran etiquetados con diversas anotaciones útiles para su análisis y clasificación, y fueron adquiridos desde 2 canales (0 y 1). En [1] se explica detalladamente la organización de esta base de datos. Se utilizó el canal 0 que, en la mayoría de los casos, corresponde a la derivación MLII. La AAMI (*Association for the Advanced of Medical Instrumentation*), sugiere una clasificación de arritmias en grupos y recomienda unos estándares para reportar resultados de desempeño de algoritmos de análisis de ECG [2]. Los grupos son 5: Ectópico supraventricular (S), ectópico ventricular (V), latidos fusionados (F), latidos desconocidos (Q) y otros latidos (N). En este trabajo se tienen en cuenta estas recomendaciones.

Los registros fueron normalizados para evitar el efecto de DC y ajustar la señal de tal forma que la máxima amplitud sea 1. La etapa de segmentación de los latidos se realizó como se sugiere en [3] y [4].

2.2. Conjunto de características

El conjunto de datos corresponde a características tomadas de trabajos previos que han reportado buen desempeño en la caracterización de la morfología, variabilidad y representación de la señal. El conjunto de características se compone de:

- Características de prematuridad: RR, post-RR, pre-RR [4].
- Características de representación: coeficientes wavelet de aproximación y detalle con función madre db2 y coeficientes de Hermite del QRS [5].
- Características morfológicas: energía y polaridad del QRS [3].

La descripción detallada de estas características se encuentra en un estudio previo [6].

En este trabajo, \mathbf{x}_i denota la i -ésima observación de q atributos o características y $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ denota la matriz de datos.

2.3. Selección de características

En [7] se presenta una definición de relevancia en términos de una matriz de *afinidad* o kernel, que captura los productos internos de las características, y un vector de ponderación. Este concepto se basa en la coherencia de los subconjuntos o *clusters* resultantes de un proceso de agrupamiento, empleando propiedades espectrales y análisis topológico derivado de la teoría de grafos [8]. A este método, los autores lo denominan $Q - \alpha$ debido a que α es el vector de ponderación y Q es la matriz ortonormal de rotación. El kernel se puede obtener con:

$$C_\alpha = \sum_{i=1}^q \alpha_i \mathbf{x}_i^\top \mathbf{x}_i = \mathbf{X} \text{Diag}(\alpha) \mathbf{X}^\top \quad (1)$$

A partir de la matriz C_α se desarrolla un método iterativo para sintonizar la matriz Q y el indicador de la relevancia α . El método se explica ampliamente en [7], [4].

2.4. Clasificación no supervisada

El agrupamiento no supervisado se realiza a través de un modelo generalizado para clustering basado en centroides (CBC), en donde se considera la proporción o grado de pertenencia de un punto a un cluster (m) y la influencia o peso de cada elemento en la actualización de los centroides (w), como se explica en [9]. El grado de pertenencia de un elemento a un cluster lo determina una función de membresía y se denota con $m(\mathbf{q}_j/\mathbf{x}_i)$ (grado de pertenencia de \mathbf{x}_i al cluster cuyo centroide es \mathbf{q}_j). Las funciones, m y w , están directamente relacionadas con la naturaleza de la función objetivo. Este modelo se basa en el principio del algoritmo H-means, es decir que se actualizan todos los centroides antes de aplicar el control de convergencia [10]. En este trabajo se aplica un algoritmo de clustering basado en densidades (DBC) obtenidos con el modelo generalizado: Clustering basado en máxima esperanza Gaussiana empleando estimación no paramétrica a través del método de Parzen (NPDBC), como se explica en [11].

La estimación del número de grupos se realiza como se sugiere en [12]. El algoritmo de clustering explicado en [8] se usa para generar la partición inicial del método NPDBC, con 2 iteraciones. Este método es de costo computacional elevado pero converge en pocas iteraciones y genera óptimos locales aproximados a los globales.

3. Resultados y discusión

El Cuadro 1 muestra los resultados generados con el método propuesto, obtenidos con 30 registros. Se muestran los valores de media y desviación estándar ($\mu-\sigma$). Las medidas de desempeño se aplicaron como se planteó en [4], para cada grupo de arritmias.

Los métodos DBC presentan buenos resultados porque usan información estadística, como el segundo momento estadístico y la probabilidad a posteriori, para el refinamiento de

la partición. Además, son menos sensibles a la inicialización que los métodos clásicos. La inicialización espectral y la selección de características mejoran el tiempo y valor de convergencia con respecto a los métodos tradicionales.

Cuadro 1. Desempeño del sistema de análisis de arritmias

Medida	N	S	V	F	Q	Tiempo (s)
Se	99.25-2.48	91.11-15.66	96.11-8.24	70.73-32.05	91.9-17.83	37.1965 - 7.6469
Sp	95.77-9.12	99.36-2.19	99.87-0.24	99.59-0.77	99.79-0.47	
Sel	99.16-2.5	96.18-6.15	98.29-3.37	93.29-10.62	99.91-0.21	

4. Conclusiones

Un método de análisis basado en clustering de tipo espectral y particional, teniendo en cuenta el costo computacional, puede lograrse de forma que los algoritmos de análisis espectral se encarguen de la inicialización, estimación del número de grupos y selección de características, y los algoritmos particionales realicen el agrupamiento de los datos.

Como trabajo futuro se pretende desarrollar opciones eficaces, para aproximarse a la solución del problema del agrupamiento de patrones homogéneos de latidos, considerando clases desbalanceadas, variabilidad morfológica y costo computacional.

Referencias

1. Moody G, Mark R. The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it. Computers in Cardiology. 1999;.
2. Recommended Practice for Testing and Reporting Performance Results of Ventricular Arrhythmia Detection Algorithms. AAMI (Association for the Advancement of Medical Instrumentation); 1998.
3. Cuesta D, Biagetti M, Micó-Tormos RQP, Aboy M. Unsupervised Detection of Ventricular Extrasystoles Using Bound- ed Clustering Algorithms and Morphology Matching. IEEE tran on Biomed. 2006;.
4. Rodríguez-Sotelo J, Cuesta-Frau D, Castellanos-Domínguez G. Unsupervised classification of atrial heartbeats using a prematurity index and wave morphology features. Medical and Biological Engineering in biomedicine. 2008;.
5. Peluffo D, Rodríguez JL, Castellanos G. Metodología para la reconstrucción y extracción de características del complejo QRS basada en el modelo paramétrico de Hermite. V Semana Técnica de Ingenierías Eléctrica y Electrónica. 2008;.
6. Sotelo JLR, Frau DC, Ordóñez DP, Domínguez GC, Novak D. Unsupervised Feature Selection in Cardiac Arrhythmia Analysis. Engineering in medicine and conference EMBC. 2009;.
7. Wolf L, Shashua A. Feature selection for unsupervised and supervised inference: The emergence of sparsity in a weight-based approach. Journal of machine learning. 2005;6:1855 – 1887.
8. Stella Y, Jianbo S. Multiclass Spectral Clustering. In: ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision. Washington, DC, USA: IEEE Computer Society; 2003. p. 313.
9. Hamerly G, Elkan C. Alternatives to the k-means algorithm that find better clusterings. In: CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management. New York, NY, USA: ACM; 2002. p. 600–607.
10. Hansen P, Mladenovic N. J-means: a new local search heuristic for minimum sum of squares clustering. Pattern Recognition. 2001;p. 405 – 413.
11. Rodríguez JL, Peluffo D, Cuesta D, Castellanos G. Non-parametric density-based clustering for cardiac arrhythmia analysis. Computers in cardiology CINC. 2009;.
12. Ng AY, Jordan MI, Weiss Y. On Spectral Clustering: Analysis and an algorithm. In: Advances in Neural Information Processing Systems 14. MIT Press; 2001. p. 849–856.