

UNIVERSIDAD POLITÉCNICA DE VALENCIA



DEPARTAMENTO DE INFORMÁTICA
DE SISTEMAS Y COMPUTADORES

TESIS DOCTORAL

NUEVOS DESARROLLOS Y APLICACIONES BASADOS EN MÉTODOS ESTOCÁSTICOS PARA EL AGRUPAMIENTO NO SUPERVISADO DE LATIDOS EN SEÑALES ELECTROCARDIOGRÁFICAS

Autor: Pau Micó
Director: David Cuesta Frau
Departamento de Informática de Sistemas y Computadores
Universidad Politécnica de Valencia

*Escola Politècnica Superior d'Alcoi
Decembre de 2005*

Abstract

The work developed and presented in this thesis continues the research line started with another PhD. thesis entitled "Estudio de métodos para procesamiento y agrupación de señales electrocardiográficas". The algorithms and methods used here have been developed to automatically process the information contained within a special kind of long-time electrocardiographic registers (called Holter ECG). The aim of the work is to provide doctors and cardiologists with a complete set of tools that make them easier the task of analysing and diagnosing the cardiac diseases. To perform this, we will apply to the Holter ECG a clustering process in order to automatically group the heart beats that compose the signal into a very few clusters from where doctors, by means of the manual inspection of a representative beat from each cluster, easily and quickly provide a diagnosis. Finally, to achieve the clustering objective, we will study the ECG morphological features using the large amount of databases available through the internet, the development of new ECG signal applications, the comparison among methods, and some algorithm optimization tasks have been also performed. A polygonal approximation algorithm used for ECG compression, the Principal Component Analysis (PCA) scheme applied to the feature selection stage or heart beat modelling by means of the Hidden Markov Models (HMM) for feature reduction too. The application of all the methods described above has become an improvement of the final Holter ECG clustering process.

Resumen

El trabajo que a continuación se presenta, pretende dar continuidad a la línea de trabajo iniciada con una tesis previa titulada "Estudio de métodos para procesamiento y agrupación de señales electrocardiográficas". La problemática a abordar es la misma que en la tesis mencionada: reducción del número de latidos a examinar en un registro electrocardiográfico (ECG) de larga duración (conocido como Holter), mediante técnicas de agrupamiento no supervisado. Sin embargo, en este caso y como técnica de extracción de características se utiliza la aproximación poligonal (Polygonal Approximation, PA) y además, en la etapa de reducción de características, se usará un método estocástico basado en los modelos ocultos de Markov (Hidden Markov Models, HMM) y uno estadístico basado en el análisis de componentes principales (Principal Component Analysis, PCA). En esta tesis se describen las aportaciones realizadas en estas tres áreas: PA, HMM y PCA, aplicadas a esta problemática concreta. La validez de los métodos viene refrendada por un conjunto amplio de experimentos utilizando como fuentes de datos las señales ECG disponibles en la base de datos del Massachusetts Institute of Technology (MIT). Como conclusión se enumeran las mejoras obtenidas de la aplicación de estos métodos sobre el proceso de clustering de latidos sobre señales ECG de tipo Holter.

Resum

El treball que a continuació es presenta vol donar continuïtat a la línia de treball iniciada amb una tesi prèvia titulada "Estudio de métodos para procesamiento y agrupación de señales electrocardiográficas". La problemàtica a tractar és la mateixa que en la tesi esmentada: reducció del nombre de batecs a examinar en un registre electrocardiogràfic (ECG) de llarga durada (conegut com a Holter), per mitjà de tècniques d'agrupament no supervisat. En canvi, en aquest cas i com a tècnica d'extracció de característiques s'empra l'aproximació poligonal (Polygonal Approximation, PA) i a més a més, en l'etapa de reducció de les característiques, s'emprarà un mètode estocàstic basat en els models ocults de Markov (Hidden Markov Models, HMM) i un altre estadístic basat en l'anàlisi de components principals (Principal Component Analysis, PCA). En aquesta tesi es descriuen les aportacions realitzades en aquestes tres àrees: PA, HMM i PCA, aplicades a aquesta problemàtica concreta. La validesa dels mètodes ve refrendada per l'ampli conjunt d'experiments realitzats emprant com a font de dades els senyals ECG de que es disposa en la base de dades del Massachusetts Institute of Technology (MIT). Com a conclusió s'enumeren les millores obtingudes de l'aplicació d'aquests mètodes sobre el procés de clustering de batecs sobre senyals ECG de tipus Holter.

Palabras clave

Procesamiento de señales biomédicas, señal electrocardiográfica, Holter, análisis de componentes principales, modelos ocultos de Markov.

Índice general

Abstract	III
Resumen	V
Resum	VII
Palabras clave	IX
I Preliminares	1
1. Introducción	3
1.1. Motivación y líneas de investigación	5
1.2. Objetivos	7
1.3. Aportaciones	8
1.4. Organización del documento	10
2. Señales Electrocardiográficas	13
2.1. Introducción	15
2.2. Origen fisiológico	15
2.2.1. Funcionamiento del corazón	15
2.2.2. Actividad eléctrica cardíaca	16
2.3. Interpretación de las ondas componentes del ECG	18
2.4. Derivaciones	21
2.5. Patologías cardíacas	23
2.6. Adquisición de la señal	23
2.6.1. Bases de datos ECG	24

2.6.2. La base de datos del MIT	26
2.7. Características y composición de las fuentes	32
3. Estado del Arte	37
3.1. Técnicas de preprocesamiento	39
3.1.1. Ruido	39
3.1.2. Interferencia de la red	48
3.1.3. Variaciones de la línea base	51
3.2. Extracción de características	56
3.3. Segmentación	57
3.3.1. Detección de las ondas <i>P</i> y <i>T</i>	57
3.3.2. Detección del complejo <i>QRS</i>	58
3.4. Reducción de características	64
3.4.1. Análisis de componentes principales	64
3.4.2. Modelos ocultos de Markov	67
3.5. Reconocimiento de patrones	92
3.5.1. Métodos supervisados	93
3.5.2. Métodos no supervisados	94
3.6. Conclusiones	105
II Caracterización de la señal electrocardiográfica	107
4. Extracción de características	109
4.1. Introducción	111
4.2. El problema de la aproximación	113
4.3. Cálculo del error	115
4.4. Definición del umbral	121
4.5. Proceso de aproximación poligonal	122
4.6. Experimentos y resultados	126
4.7. Conclusiones	133
5. Preprocesamiento y segmentación	139
5.1. Introducción	141
5.2. Interferencia de la red eléctrica	141
5.3. Variaciones de la línea de base	142
5.4. Ruido interferente	143

5.5. Segmentación de la señal en latidos	144
5.6. Normalización de los latidos	147
5.7. Conclusiones	148
III Clustering de latidos	149
6. Introducción al clustering	151
6.1. Fases del proceso	153
6.2. Selección de características	153
6.3. Medidas de disimilitud	154
6.4. Criterios de clustering	155
6.5. Algoritmos de clustering utilizados	155
6.5.1. Algoritmo k-medias	156
6.5.2. Algoritmo Max-Min	158
6.5.3. Clustering jerárquico binario	160
6.6. Selección del número óptimo de clusters	163
6.7. El problema de los outliers	165
6.8. Diagnóstico y validación de los clusters	167
6.9. Conclusiones	170
7. Clustering mediante Análisis de Componentes Principales	171
7.1. Generalidades sobre las pruebas con PCA	173
7.2. Generación de la matriz de datos	173
7.3. Identificación de los outliers	175
7.4. Optimización del número de clusters	177
7.5. Optimización para el clustering K-medias	178
7.6. Optimización para el clustering Max-Min	179
7.7. Optimización para el clustering binario	180
7.8. Comparativa entre los algoritmos de clustering utilizados	181
7.9. Conclusiones	182
8. Clustering utilizando Modelos Ocultos de Markov	187
8.1. Introducción	189
8.2. El problema del clustering con HMM	190
8.3. Descripción del proceso de clustering con HMM	190
8.3.1. Adquisición y preprocesado	191

8.3.2. Selección de los centroides iniciales	193
8.3.3. Clustering de latidos	195
8.3.4. Entrenamiento	197
8.3.5. Consolidación y evaluación	199
8.4. Experimentos y resultados	201
8.4.1. Experimentos previos al clustering	201
8.4.2. Experimentos de clustering	209
8.5. Conclusiones	213
IV Conclusiones	221
9. Conclusiones y futuras líneas de investigación	223
9.1. Conclusiones	225
9.2. Futuras líneas de investigación	231
Bibliografía	235
Lista de acrónimos	257

Índice de cuadros

2.1. Duración del intervalo QT	21
2.2. Descripción de las principales derivaciones	23
2.3. Derivaciones ECG y polaridad de las ondas	24
2.4. Composición de la base de datos QT	30
2.5. Características de los registros electrocardiográficos (i)	33
2.6. Características de los registros electrocardiográficos (ii)	34
2.7. Composición en latidos del ECG (i)	34
2.8. Composición en latidos del ECG (ii)	35
3.1. Comparativa entre algunos filtros de ventana	43
3.2. Algoritmos de detección del complejo QRS	63
4.1. Umbrales de error de la aproximación poligonal	130
8.1. Optimización del número de estados (ii)	208
8.2. Mejora obtenida con el entrenamiento de los modelos	212
8.3. Clustering sobre una colección de latidos de gran variabilidad	213

Índice de figuras

1.1. Proceso general de clustering de latidos	7
2.1. Fisiología del músculo cardíaco	16
2.2. Actividad eléctrica cardíaca	18
2.3. Modelo ideal del ciclo cardíaco	19
2.4. Espectro en potencia de la señal electrocardiográfica	22
2.5. Disposición de los electrodos en el triángulo de <i>Einthoven</i>	22
2.6. Aspecto de la web de PhysioBank	28
3.1. Señal ECG con ruido	40
3.2. Sistema básico de filtrado	41
3.3. Espectro de una señal electrocardiográfica real	42
3.4. Eliminación del ruido mediante aproximación poligonal	45
3.5. Señal ECG con interferencia de la red	49
3.6. Estructura del filtro para eliminar la interferencia de la red	51
3.7. Señal ECG con variaciones en la línea base	52
3.8. Extracción de características mediante aproximación poligonal	57
3.9. Detección de las ondas <i>P</i> y <i>T</i>	59
3.10. Detección del complejo <i>QRS</i>	59
3.11. Cadena de Markov de dos estados	71
3.12. HMM discreto de tres estados	73
3.13. Diferencias entre un HMM y un FHMM	82
3.14. Arquitectura de un HMM de entrada-salida	83
3.15. Árboles de decisión de Markov	84
3.16. Arquitectura híbrida HMM/NN	85
3.17. HMM auto-regresivo	85

3.18. Modelos de Markov enterrados	86
3.19. HMM acoplado	87
3.20. HMM de duración variable	88
3.21. HMM segmentado	89
3.22. Ejemplo de clustering particional	100
3.23. Ejemplo de clustering jerárquico	100
4.1. Ejemplo de una aproximación poligonal	115
4.2. Estimación del umbral de error	116
4.3. Error absoluto	117
4.4. Error de superficie	119
4.5. Error basado en la distancia Euclídea modificada	120
4.6. Obtención del Error Umbral	123
4.7. Efecto de filtrado paso bajo	124
4.8. Tiempo de procesamiento en el cálculo del umbral de error . . .	127
4.9. Tiempo de procesamiento de la aproximación poligonal	128
4.10. Aproximación poligonal con relajamiento del umbral	129
4.11. MAE para las diferentes métricas	130
4.12. MSE para las diferentes métricas	131
4.13. CRP para las distintas aproximaciones	132
4.14. Velocidad de procesamiento en señales comprimidas	133
4.15. Mejora en el procesamiento respecto al error cuadrático	135
5.1. Aproximaciones para la corrección de la línea base	142
5.2. Mejor nivel para la aproximación de la línea de base	143
5.3. ECG con la línea de base corregida	144
5.4. Eliminación de la línea de base	144
5.5. Detección de los puntos característicos	145
5.6. Aplicación desarrollada para el etiquetado de los latidos	146
5.7. Segmentación de latidos	146
5.8. Normalización de los latidos	147
6.1. Etapas del proceso de clustering	153
6.2. Dendrograma generado por la función criterio	160
6.3. ECG con outliers	166
6.4. Problema del clustering con outliers	167

6.5. Formas de onda de las clases	168
7.1. Selección de la propiedad en la matriz de datos	174
7.2. Identificación de los outliers con asociatividad 1	175
7.3. Identificación de los outliers con asociatividad 2	176
7.4. Selección del número de clusters óptimo	177
7.5. Precisión según el número de clusters	178
7.6. Selección de la métrica óptima para el algoritmo K-medias . . .	179
7.7. Selección de la métrica óptima para el algoritmo Max-Min . . .	180
7.8. Selección de la métrica óptima para el algoritmo Binario	181
7.9. Comparativa entre los diferentes algoritmos de clustering	182
8.1. Diagrama de bloques del clustering con HMM	192
8.2. Latidos caracterizados con 25 segmentos	202
8.3. Selección de la dimensión y tipo de las características	203
8.4. Gráficos de dispersión de las matrices de probabilidad	205
8.5. Selección de los centroides iniciales	206
8.6. Optimización del número de estados (i)	207
8.7. Optimización del número de estados (ii)	208
8.8. Clustering con optimización de modelos (i)	210
8.9. Clustering con optimización de modelos (ii)	211
8.10. Reducción de características sobre una señal ECG real	211
8.11. Clustering de latidos sobre una señal ECG real	212
8.12. Clustering de latidos: modelos iniciales vs entrenados	218

Parte I

Preliminares

Capítulo 1

Introducción

***Resumen.** En este capítulo se establece el contexto en el que se va a desarrollar el presente trabajo de investigación, detallándose los motivos que dieron lugar al mismo. A continuación se enumeran tanto los objetivos perseguidos como las aportaciones obtenidas como resultado de las aplicaciones desarrolladas en cada uno de los capítulos que componen esta tesis doctoral. Por último, se realiza un breve resumen del contenido de los diferentes capítulos con el fin de ofrecer al lector una visión global de todo el trabajo realizado.*

Contenidos

1.1. Motivación y líneas de investigación	5
1.2. Objetivos	7
1.3. Aportaciones	8
1.4. Organización del documento	10

1.1. Motivación y líneas de investigación

Motivación

Es bien conocida la importancia de la señal electrocardiográfica (ECG) para el diagnóstico de multitud de enfermedades cardíacas, tanto mediante inspección visual como mediante técnicas de inspección automática. Al igual que en el análisis cualquier otra señal, la calidad y cantidad de información obtenida dependerá de propiedades tales como la relación señal-ruido (SNR), la resolución del conversor analógico-digital, la frecuencia de muestreo, etc. Por esta razón y previamente a su tratamiento automatizado, resulta extremadamente importante que la señal nos proporcione una información con la calidad adecuada, máxime si se tiene en cuenta que el diagnóstico de ciertas enfermedades cardiovasculares graves depende de ello. Por otro lado, en la adquisición de registros ECG de larga duración (llamados *Holter*), la cantidad de latidos a examinar por parte del equipo médico puede llegar a ser de varios miles, lo cual convierte la inspección visual de una señal de este tipo en una tarea laboriosa y compleja. De hecho, los *Holter*, corresponden a registros ambulatorios de 24 ó 48 horas, que posteriormente se transmiten al equipo de análisis a una velocidad de 60 a 120 veces la velocidad de adquisición. Después se procede a la extracción de los eventos significativos, muchas veces de forma manual. El análisis automático de estas señales constituye un interesante campo de investigación, por la dificultad que entraña su análisis manual. Sin embargo, los métodos utilizados hasta este momento para llevar a cabo esta tarea, no ofrecen resultados lo suficientemente satisfactorios y resulta necesario el revisar los resultados generados automáticamente para evitar la pérdida de latidos con información importante. En nuestro caso, pretendemos analizar distintas técnicas de agrupamiento automático (clustering) aplicadas sobre los latidos de las señales ECG *Holter* para obtener un agrupamiento por tipos de los diferentes latidos que aparecen en una señal electrocardiográfica. Cada uno de los grupos obtenidos deberá ser completamente homogéneo, de manera que únicamente contenga latidos idénticos desde el punto de vista del diagnóstico, con lo que el cardiólogo sólo deberá examinar un latido de cada grupo para determinar la patología del paciente, reduciendo significativamente la cantidad de datos a inspeccionar.

Líneas de investigación

Esta tesis pretende dar continuidad a la línea de investigación abierta por el Dr. Ing. David Cuesta Frau y que se centra en el estudio de las técnicas de tratamiento automático de señales electrocardiográficas de larga duración [Cue01]. En el documento original el Dr. Ing. David Cuesta revisa el estado del arte describiendo toda una amplia gama de técnicas de preprocesamiento de la señal ECG, desde la eliminación del ruido (mediante técnicas clásicas de filtrado, aproximación de funciones, o la transformada *wavelet*), pasando por la eliminación de la interferencia de la red, la eliminación de las variaciones de la línea base (utilizando diferentes tipos de filtros), la detección de los puntos significativos de una onda ECG (mediante algoritmos de tratamiento digital de las señales, detectando la primera y segunda derivadas, filtrado digital, transformaciones no lineales, etc.), todo ello para conseguir una señal ECG limpia de interferencias y perfectamente segmentada por latidos. Además, dedica todo un capítulo a las operaciones de preprocesamiento basadas en la transformada *wavelet*. A partir de este punto y tomando como fuente los latidos previamente segmentados, se presentan una serie de métodos para el procesado (normalización temporal lineal, alineamiento temporal no lineal, variaciones para la extracción de características) y clustering final de los latidos. Finalmente el trabajo concluye presentado los resultados obtenidos de la aplicación práctica de todos los métodos y algoritmos descritos.

En la tesis que a continuación se desarrolla, y siguiendo con los trabajos anteriormente mencionados, se propone la mejora del proceso de clustering de latidos mediante la aplicación de diversas técnicas en alguna de las etapas de dicho proceso (*Figura 1.1*). Las aportaciones realizadas por el autor van encaminadas en dos direcciones, pretendiendo optimizar por un lado el proceso de extracción de características y, por otro lado, mejorando la etapa de clustering de latidos mediante la aplicación de técnicas novedosas. En concreto, el autor propone una etapa de extracción de características mediante aproximación poligonal de los latidos inmediatamente después (o en paralelo) al proceso de adquisición de la señal. El beneficio derivado de una extracción previa al procesamiento del ECG supone, además de la compresión de la información almacenada, una disminución del tiempo de procesamiento en tareas tales como la eliminación de la línea base, filtrado del ruido, etc. Es éste el motivo

por el cual se procede a la extracción de las características *antes* del preprocesamiento de la señal. Además, y en la etapa de selección o extracción de las características se pretende la utilización de dos métodos: uno estadístico como el (i) *Análisis de Componentes Principales* (Principal Component Analysis, PCA) y un método estocástico de modelado como el de los (ii) *Modelos Ocultos de Markov* (Hidden Markov Models, HMM), que también se utilizará durante el proceso de clustering.

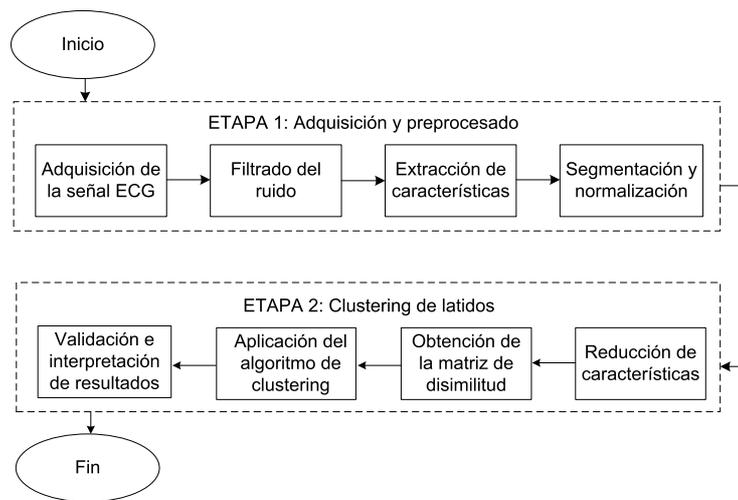


Figura 1.1: Proceso general desarrollado para el clustering de latidos de una señal ECG de tipo Holter.

1.2. Objetivos

Los objetivos perseguidos, y siempre teniendo en cuenta que el origen de datos va a ser una señal biomédica de tipo electrocardiográfico, son los que se detallan a continuación:

1. Estudio de las características de la señal electrocardiográfica así como de las diferentes ondas que la componen.

2. Recopilación de una colección de señales electrocardiográficas lo suficientemente variadas y representativas como para que nos permitan llevar a cabo un estudio experimental.
3. Estudio y evaluación de las diferentes técnicas existentes en la bibliografía para el preprocesamiento de señales electrocardiográficas.
4. Estudio comparativo de los métodos de aproximación poligonal existentes para el tipo de señales estudiado y su posterior aplicación en la optimización del proceso de compresión de la señal ECG.
5. Estudio del PCA como técnica de selección de características de los latidos. Validación del método mediante el análisis de los resultados obtenidos tras su aplicación en el proceso de clustering de señales ECG reales de tipo Holter.
6. Estudio de los HMMs como técnica de selección de las características de los latidos mediante modelado estadístico. Análisis de los resultados y validación del método tras su aplicación en el proceso de clustering de señales ECG reales de tipo Holter.

Con todo esto podemos resumir el objetivo final de la tesis como el de la optimización de los resultados derivados del clustering de latidos mediante la aplicación de diferentes métodos de mejora en varias de las etapas que componen el proceso.

1.3. Aportaciones

El trabajo desarrollado en esta tesis ha supuesto la presentación de numerosas ponencias y comunicaciones en distintos congresos internacionales y la elaboración de varios artículos de revista que se encuentra actualmente en proceso de revisión. En cada uno de los capítulos donde se han implementado aplicaciones prácticas, el autor ha validado los resultados derivados de la aplicación de las diferentes técnicas de análisis anteriormente citadas, al menos en un congreso internacional. Estas publicaciones son las siguientes:

Capítulo 4: Extracción de características. De los resultados obtenidos en la compresión del ECG se ha presentado en congreso internacional un artículo donde se describe la extracción de características mediante la aproximación poligonal de la señal ECG así como la optimización del proceso de aproximación tras el análisis comparativo de los resultados derivados de la aplicación de métricas diferentes [MCN05b].

Capítulo 5: Preparación de la señal ECG para el clustering. Justificando el proceso de preparación de la señal ECG se han presentado varios artículos en congresos internacionales. En [MC02] se muestran algunas de las técnicas de preprocesamiento de señales biomédicas en general que luego son implementadas y aplicadas a las señales ECG en [CMA⁺03].

Capítulo 7: Análisis de componentes principales. Donde se evalúa la aplicación de PCA en la caracterización de los latidos para su posterior clustering. La utilización de esta técnica así como los resultados obtenidos se han contrastado mediante la publicación en congreso internacional del correspondiente artículo [MCN05a]. Además, en [MCN04a] se implementa y evalúa una aplicación del método de PCA para la optimización del tiempo de procesamiento en la etapa de selección de características de los latidos del Holter.

Capítulo 8: Modelos ocultos de Markov. En cuanto a los artículos generados en base al clustering de latidos donde se propone como técnica de reducción de características el modelado mediante *Modelos Ocultos de Markov*, cabe destacar aquellos en los que se describe el proceso global de análisis y clustering de señales biomédicas con HMM [CM02] y [NCML04]. Además, también se han tratado de optimizar los HMMs (sobre todo en la etapa de su inicialización o en la de su entrenamiento) como en [NCML03] donde se desarrolla un método de estimación del número de clusters; o en [NAaH⁺04] donde se propone la técnica de *simulated annealing* para la mejora del proceso de entrenamiento y aprendizaje de los modelos. En [MCN04b] se sugiere una nueva forma de inicializar los HMM en una etapa de preclustering. Por otra parte, y para complementar los desarrollos presentados en este capítulo, también se han realizado estudios comparativos para la aplicación al clustering de diferentes métodos como el de Dynamic Time Warping (DTW) y el anteriormente

citado de HMM [NCAa⁺04]. Finalmente en [CMNA04] se propone la aplicación al procesamiento de señales biomédicas de diferentes métodos de comparación de patrones (pattern matching), entre los que destaca el de HMM.

Actualmente se encuentra en proceso de revisión un artículo de revista (IEEE) titulado *Clustering of Electrocardiogram Signals using Hidden Markov Models* en el que se describen algunos de los métodos aplicados en la mejora del modelado de latidos mediante HMM, así como el resultado final del proceso de clustering [NCAa⁺05].

Además, también se encuentra en proceso de elaboración un artículo de revista titulado *Clustering of Electrocardiographic Signals* donde se tratan de resumir todos los desarrollos y resultados de clustering obtenidos derivados de la presente tesis doctoral.

1.4. Organización del documento

El presente trabajo está organizado en cuatro partes cuyo contenido se detalla a continuación:

Parte I. Dedicada a los preliminares donde se describen las principales características fisiológicas de la señal ECG y se analiza el estado del arte. Está formada por los siguientes capítulos:

- **Capítulo 1.** A modo de introducción, donde se detallan los objetivos a alcanzar, se enumeran las aportaciones realizadas y se describe el contenido del documento por capítulos.
- **Capítulo 2.** En el que se describen las características más importantes de las señales electrocardiográficas, proporcionando la información necesaria para una primera interpretación de los resultados del clustering de latidos. También se detalla el origen de las fuentes ECG utilizadas en el trabajo así como su composición en latidos.
- **Capítulo 3.** Donde se analiza el actual estado del arte de técnicas de preprocesamiento de la señal ECG, análisis mediante PCA y estado de desarrollo y aplicación de los HMM.

Parte II. En la que se describen la etapa de extracción de las características y todas las tareas de preparación de la señal ECG previas al clustering de latidos.

- **Capítulo 4.** Extracción de las características de la señal ECG mediante el estudio y optimización del algoritmo de aproximación poligonal aplicado sobre el Holter.
- **Capítulo 5.** Donde se analizan las operaciones de preprocesamiento que sufre la señal ECG durante la preparación de los latidos antes de iniciar el proceso de clustering. Además, también se desarrolla el proceso de segmentación que nos permitirá identificar y separar los latidos del ECG.

Parte III. En ella se describe el proceso de clustering, particularizando en la etapa de selección de las características mediante PCA y por modelado con HMM.

- **Capítulo 6.** Este capítulo está dedicado a la descripción general de un proceso de clustering, sin particularizar en su aplicación sobre ningún tipo de señal en concreto.
- **Capítulo 7.** Donde se utiliza el método de PCA para la extracción de características de los objetos a agrupar. Al final del capítulo se proporcionan los resultados y conclusiones derivados del clustering con reducción por PCA.
- **Capítulo 8.** Se desarrolla un modelado de los latidos con HMM reduciendo el número de características para la posterior aplicación de un algoritmo de clustering basado en el refinamiento de los modelos entrenados.

Parte IV. Compuesta por un único capítulo en el que se aglutinan las conclusiones obtenidas de cada uno de los capítulos anteriores así como las futuras líneas de investigación.

- **Capítulo 9.** Capítulo donde se presentan las conclusiones derivadas de la aplicación de diferentes métodos de caracterización y modelado de latidos en el proceso de clustering de latidos de una señal Holter ECG. Además se proponen las futuras líneas de investigación.

Capítulo 2

Señales Electrocardiográficas

Resumen. En este capítulo se describen desde un punto de vista genérico las características más importantes de la señal electrocardiográfica o electrocardiograma (ECG), sus ondas componentes y sus intervalos. Resulta evidente que para un estudio completo de las señales electrocardiográficas se requiere un nivel de detalle mucho mayor [Har90], objetivo que queda fuera del ámbito de este trabajo. En nuestro caso, sólo necesitamos de un conocimiento básico de las principales formas de onda que en ella aparecen para poder interpretar los resultados que se expondrán en capítulos posteriores. Además, también se detalla el método de adquisición de la señal ECG, así como origen de las señales utilizadas en el trabajo y su composición en latidos.

Contenidos

2.1. Introducción	15
2.2. Origen fisiológico	15
2.2.1. Funcionamiento del corazón	15
2.2.2. Actividad eléctrica cardíaca	16
2.3. Interpretación de las ondas componentes del ECG	18
2.4. Derivaciones	21
2.5. Patologías cardíacas	23
2.6. Adquisición de la señal	23
2.6.1. Bases de datos ECG	24
2.6.2. La base de datos del MIT	26
2.7. Características y composición de las fuentes	32

2.1. Introducción

A continuación se describen las principales características de la señal ECG que nos ayudarán a comprender las razones que nos han llevado a utilizar los métodos y algoritmos presentados en capítulos posteriores. El *Holter* (o señal de electrocardiografía dinámica ambulatoria) corresponde a un registro ECG continuo de 24 ó 48 horas que permite un mayor control de la evolución de un paciente a lo largo de un período relativamente grande de tiempo, evitando la falta de información que aparece en un ECG normal debido a su corta duración. Se ha demostrado su utilidad en la documentación de arritmias y trastornos de conducción del músculo cardíaco. Mediante la correlación de los síntomas de los pacientes y los registros electrocardiográficos puede hallarse una explicación a síntomas como mareos, síncope y palpitaciones [Lag90]. Un *Holter* está formado por centenares de miles de ondas, cada una correspondiente a un latido, lo cual hace imposible su examen detallado mediante inspección visual. Para evitar esto, la solución adoptada habitualmente es la de agrupar las ondas pertenecientes a un mismo tipo, examinando únicamente una señal representativa de cada clase, por lo que será conveniente aplicar un tratamiento que automatice la tarea de clasificación (*Capítulo 3*).

2.2. Origen fisiológico

El origen fisiológico de la señal electrocardiográfica se puede resumir en dos preguntas básicas: ¿Cómo funciona el corazón? y ¿Cuál es su actividad eléctrica?

2.2.1. Funcionamiento del corazón

El corazón humano (*Figura 2.1*) es en realidad el resultado de la unión de dos corazones: el derecho que envía la sangre sin oxígeno al pulmón para que se oxigene; y el izquierdo, que envía la sangre oxigenada al resto del cuerpo. Cada corazón tiene a su vez dos cavidades: las aurículas derecha e izquierda que serían la antesala; y los ventrículos derecho e izquierdo actuando como una bomba sanguínea que con su potente contracción muscular impulsa la

sangre hacia el pulmón o hacia el resto del cuerpo. La contracción del ventrículo derecho es de menor fuerza que la del izquierdo ya que al primero le basta una pequeña contracción para impulsar la sangre al pulmón (se podría decir que la aurícula y el ventrículo derecho son de baja presión) mientras que el ventrículo izquierdo tiene que impulsar la sangre a gran presión para que esta llegue a todos los órganos incluido el sistema nervioso central (alta presión). Las aurículas están separadas de los ventrículos por las válvulas auriculoventriculares: la tricúspide en el lado derecho y la mitral en el izquierdo. Ambas válvulas impiden que la sangre de los ventrículos retroceda a las aurículas cada vez que éstos se contraen. Ambos corazones (figuradamente el derecho y el izquierdo) están totalmente separados por dos tabiques: el interatrial (que separa a las aurículas) y el interventricular (que separa a los ventrículos) de forma que la sangre no oxigenada no se mezcle con la sangre oxigenada.

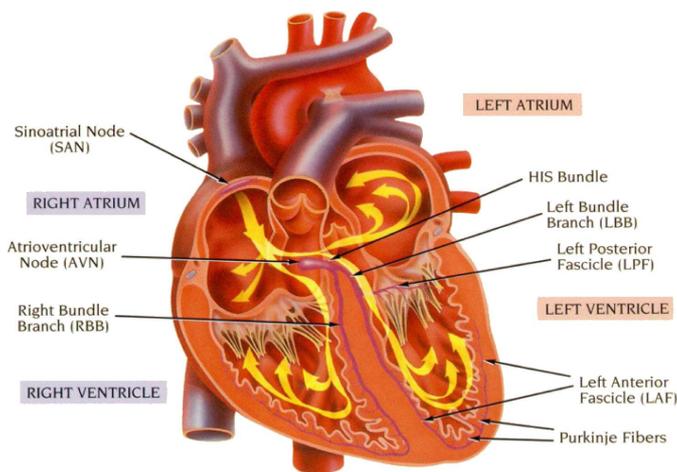


Figura 2.1: Fisiología del músculo cardíaco

2.2.2. Actividad eléctrica cardíaca

El electrocardiograma es una exploración básica en cardiología que consiste en la obtención, mediante unos electrodos aplicados en la piel, de un registro con la actividad eléctrica del corazón [Lag90]. El registro electrocar-

diográfico consiste en una línea de base y varias deflexiones y ondas. Mediante el análisis de este registro se pueden detectar crecimientos de las cavidades cardíacas y alteraciones del ritmo del corazón, empleándose de manera rutinaria en el diagnóstico inicial de muchas enfermedades cardíacas y constituyendo la base de otras exploraciones más complejas como la electrocardiografía de esfuerzo (prueba de esfuerzo) o la electrocardiografía dinámica ambulatoria (Holter).

Primeramente, y para conocer con detalle las principales características de la señal electrocardiográfica, pasaremos a describir la generación de los impulsos eléctricos que originan el registro electrocardiográfico. El origen de este impulso se encuentra en las células del músculo cardíaco, las cuales pueden ser excitadas eléctricamente, produciéndose un trasiego de iones a través de su membrana, lo cual induce un potencial eléctrico variable en el interior y en el exterior. El corazón se contrae únicamente si el músculo cardíaco cambia la conductividad de su pared celular, permitiendo el flujo de iones de calcio como elemento más característico dentro de un complejo proceso. En general, la contracción es disparada desde un pequeño grupo de células conocido como nodo sinusal o nódulo de Keith-Flach, que se despolarizan automáticamente cada 800 ms. Este nodo se encuentra localizado en la parte superior de la aurícula derecha (posición cefálica), concretamente en la desembocadura de la vena cava superior. El nodo sinusal es el principal marcapasos del corazón por su capacidad de producir un mayor número de despolarizaciones por minuto (60-80 latidos/minuto). La despolarización se transmite a las células auriculares adyacentes, describiendo un flujo de cationes Na^+ y Ca^{++} , que irrumpen desde el medio extracelular hacia el interior cuando la membrana se hace permeable. De esta forma, el estímulo se propaga por todo el miocardio auricular produciendo su contracción. La fase de recuperación se conoce como repolarización y durante este período las concentraciones de iones vuelven a su nivel normal. En reposo las células tienen un potencial transmembrana de 90 mV. Este potencial es cíclico, con un período que oscila entre los 400 y los 1200 ms. Posteriormente, el estímulo generado con la polarización auricular alcanza la unión auriculoventricular (nodo AV), que está a su vez conformada por tejido automático (nodo de Aschoff-Tawara) y por tejido de conducción (haz de His) avanzando de aurícula a ventrículo, evitando frecuencias cardíacas superiores a 200 lpm (latidos/minuto) [KJ94]. Desde el nodo AV surgen

dos ramas, la izquierda y la derecha, por donde el estímulo eléctrico se distribuye por ambos ventrículos a través del sistema de Purkinje polarizando los ventrículos y provocando de esta forma la contracción ventricular.

La despolarización del músculo produce un nivel positivo que precede a la onda de avance. Esto significa que en la superficie del músculo los electrodos recogen un nivel positivo. La amplitud de la deflexión es proporcional a la masa del músculo, lo cual permite detectar, por ejemplo, a pacientes con los ventrículos derecho o izquierdo agrandados (hipertrofia ventricular). Los transductores encargados de medir la diferencia de potencial en la superficie del cuerpo son los electrodos. En la *Figura 2.2* se muestra la estructura genérica del sistema nervioso del corazón.

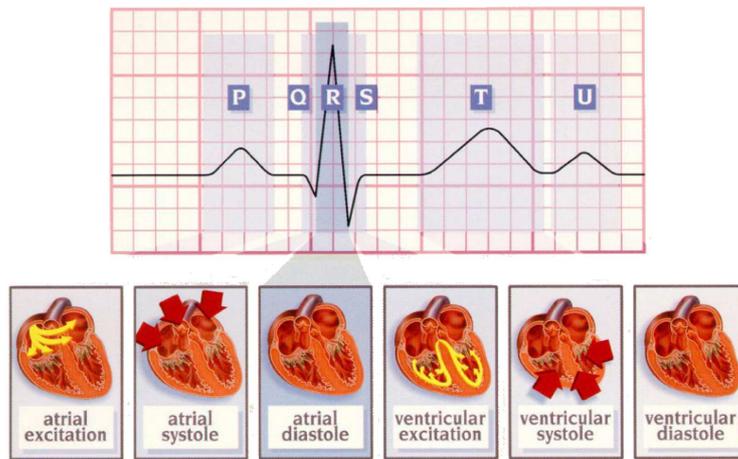


Figura 2.2: Interpretación de la actividad eléctrica cardíaca, donde relaciona la actividad muscular del corazón con la generación de la onda electrocardiográfica

2.3. Interpretación de las ondas componentes del ECG

Durante las fases de despolarización y repolarización miocárdica y sobre el ECG registrado aparecen una serie de ondas electrocardiográficas. Las dis-

tancias entre las ondas se denominan intervalos mientras que las porciones de ECG que aparecen entre las deflexiones se denominan segmentos. Un período del ECG perteneciente a un individuo sano consta de una onda *P*, un complejo *QRS*, una onda *T* y una onda *U*, tal y como se muestra en la *Figura 2.3*.

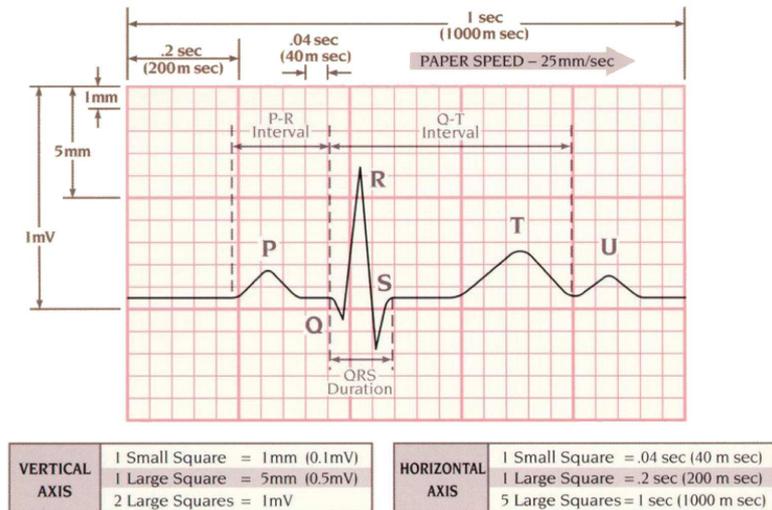


Figura 2.3: Modelo ideal del ciclo cardíaco

Dentro del ECG de un ciclo cardíaco podemos distinguir las siguientes ondas, deflexiones y segmentos:

- *Onda P*. Representa la despolarización de la aurícula (lo que supone la contracción o sístole auricular). Su duración normal es de 0,1 s. La forma de esta onda depende de la localización de los electrodos (derivación). Un aumento del voltaje de esta onda indica una anomalía auricular. La ausencia de esta onda ocurre en una parada del nodo sinusal y en el bloqueo SA sinoauricular (situación en la que sí se despolariza el nodo sinusal pero no se transmite el impulso a las células de la aurícula contiguas).
- *Complejo QRS*. Representa la despolarización de los ventrículos (lo que supone la contracción o sístole ventricular). Está formado por las ondas

Q , R y S , aunque pueden aparecer otras variantes llamadas R' y S' . Su duración aproximada es de 100 ms.

- *Onda T*. Supone la repolarización de los ventrículos (relajación o diástole ventricular). La onda T normal es asimétrica en sus ramas y está redondeada en su vértice. La pendiente de la rama inicial es más suave que la de la rama terminal. Las anomalías en esta onda pueden indicar enfermedades cardíacas primarias (aunque no necesariamente). También puede producir trastornos del equilibrio hidroelectrolítico.
- *Onda U*. Su origen fisiológico es poco claro, aunque puede ser debida a la repolarización del sistema de Purkinje. Presenta una forma anormal cuando se dan trastornos del potasio.
- *Segmento PR*. Corresponde a la línea isoeletrica definida desde el comienzo de la onda P (inicio de la despolarización auricular) hasta la deflexión inicial del complejo QRS (inicio de la despolarización de los ventrículos). La duración normal de este segmento está entre los 0,12 y los 0,21 s, variación que depende de la frecuencia cardiaca.
- *Segmento ST*. Es el intervalo entre el final del complejo QRS y el inicio de la onda T . Representa el tiempo durante el que los ventrículos permanecen en estado activado y puede iniciarse la repolarización ventricular. Normalmente el segmento ST es isoeletrico, aunque también puede aparecer ligeramente desviado. Una desviación elevada a menudo representa un infarto de miocardio, una pericarditis aguda o una miocarditis.
- *Intervalo PP*. Correspondiente al tiempo transcurrido entre el inicio de la onda P y el inicio de la siguiente onda P .
- *Intervalo RR*. Correspondiente al tiempo transcurrido entre la onda R de un complejo QRS y la onda R del siguiente complejo QRS .
- *Intervalo QRS*. Tiempo transcurrido entre el comienzo de una onda Q y el final de la onda S dentro del mismo complejo QRS . Es un indicador del tiempo de conducción intraventricular. El límite superior de duración normal del complejo QRS es inferior a los 0,12 s. Una duración superior a 0,12 s significa que el impulso se inició desde el nodo AV

(auriculo-ventricular) o incluso más arriba (impulso supraventricular). Un complejo *QRS* ancho (con una duración superior a 0,12 s) puede indicar que la conducción procede del ventrículo o del tejido supraventricular, pero que hay una conducción prolongada a través del ventrículo que genera por lo tanto un complejo *QRS* de mayor duración.

- *Intervalo QT*. Correspondiente al intervalo de tiempo transcurrido entre el comienzo del complejo *QRS* y el final de la onda *T*, representando de esta forma la duración de la sístole eléctrica. La relación entre el ritmo cardíaco y la duración de este intervalo viene dado en la *Tabla 2.1*.

RITMO CARDÍACO (LPM)	DURACIÓN QT (S)
60	0,33–0,43
70	0,31–0,41
80	0,29–0,38
90	0,28–0,36
100	0,27–0,35
120	0,25–0,32

Cuadro 2.1: Duración del intervalo QT

Además, y desde el punto de vista del procesado de señales, también hay que tener en cuenta el contenido frecuencial del ECG. El ancho de banda de la señal electrocardiográfica es aproximadamente de unos 125 Hz, con la distribución espectral mostrada en la *Figura 2.4*.

2.4. Derivaciones

Dependiendo de la posición sobre la superficie del cuerpo del paciente donde se coloquen los electrodos, se obtendrán distintas derivaciones o pares de puntos. Las derivaciones pueden ser:

- *Bipolares*. Son las que se obtienen con el llamado triángulo de *Einthoven* (*Figura 2.5*), situando los electrodos en las extremidades.
- *Aumentadas*. Una variante de la anterior.

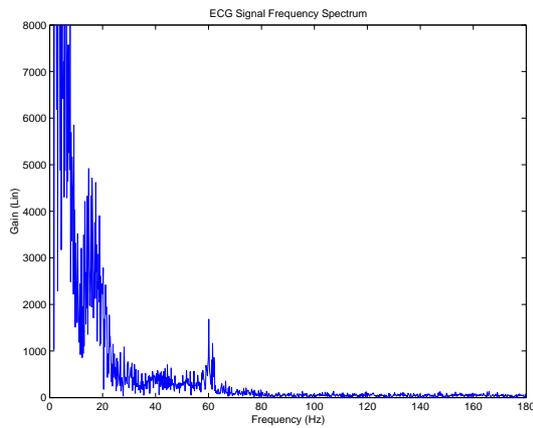


Figura 2.4: Espectro en potencia de la señal electrocardiográfica

- *Unipolares*. Se obtienen con los electrodos situados sobre el tórax del paciente.

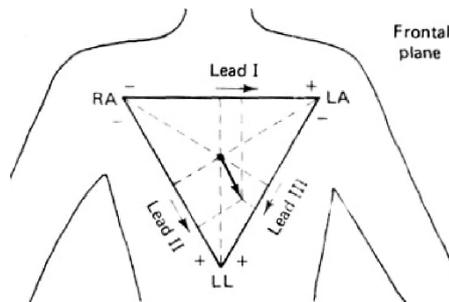


Figura 2.5: Disposición de los electrodos en el triángulo de *Einthoven*

Existen otras derivaciones menos conocidas (pero que a veces resultan importantes) que son variantes de las derivaciones unipolares: derivaciones V_7 , V_8 y V_9 que enfrentan la parte posterior del corazón; y las derivaciones RV_1 y RV_2 , que enfrentan el ventrículo derecho.

En la *Tabla 2.2* y en la *Tabla 2.3* se muestran las principales derivaciones así como la posición de los electrodos, la nomenclatura utilizada y la polaridad de las ondas registradas.

DERIVACIÓN	POSICIÓN ELECTRODOS	NOMENCLATURA
Bipolar	RL, LL, RA, LA	$I = LA - RA$ $II = LL - RA$ $III = LL - LA$
Aumentadas	RL, LL, RA, LA	$aVR = RA - 0,5(LA + LL)$ $aVL = LA - 0,5(LL + RA)$ $aVF = LL - 0,5(LA + RA)$
Unipolares	$v1, v2, v3, v4, v5, v6$	$V1 = v1 - \frac{LA+RA+LL}{3}$ $V2 = v2 - \frac{LA+RA+LL}{3}$ $V3 = v3 - \frac{LA+RA+LL}{3}$ $V4 = v4 - \frac{LA+RA+LL}{3}$ $V5 = v5 - \frac{LA+RA+LL}{3}$ $V6 = v6 - \frac{LA+RA+LL}{3}$
Ortogonales	I, E, C, A, M, H, F	$X = 0,610A + 0,171C - 0,781I$ $Y = 0,655F + 0,345M - 1,000H$ $Z = 0,133A + 0,736M - 0,264I - 0,374E - 0,231C$

Cuadro 2.2: Descripción de las principales derivaciones

2.5. Patologías cardíacas

A partir de la descripción de los componentes (segmentos y ondas) de un latido normal de la señal ECG se pueden empezar a identificar, describir y clasificar las diferentes cardiopatías que se manifiestan en el paciente con el cambio de la forma de onda calificada como *normal* y que ha sido descrita anteriormente. El estudio de cada una de las patologías, así como la justificación fisiológica de la forma de onda relacionada con cada una de ellas y su correspondiente diagnóstico es el campo de trabajo de los cardiólogos y queda fuera del ámbito de esta tesis. Un estudio completo y la clasificación de las patologías cardíacas más comunes, así como sus manifestaciones en la señal electrocardiográfica aparecen debidamente detallados en [Har90] [Lag90] y [Par99].

2.6. Adquisición de la señal

Como en cualquier sistema de procesamiento de señales, la primera etapa consiste en la adquisición de la propia señal. En la práctica, esta adquisición se realiza mediante los dispositivos tipo *Holter*[SCSA04a]. Para poder validar el

DERIVACIÓN	ONDA P	COMPLEJO QRS	ONDA T
<i>I</i>	+	+	+
<i>II</i>	+	+	+
<i>III</i>	+/-	+/-	+/-
<i>VR</i>	-	-	-
<i>aVL</i>	+/-	+/-	+
<i>aVF</i>	+	+	+
<i>V1</i>	+/-	-	-
<i>V2</i>	+/-	+/-	+/-
<i>V3</i>	+	+	+
<i>V4</i>	+	+	+
<i>V5</i>	+	+	+
<i>V6</i>	+	+	+

Cuadro 2.3: Derivaciones ECG y polaridad de las ondas

procesamiento automatizado de señales ECG es necesaria la aplicación del tratamiento sobre un conjunto más o menos amplio de señales ECG las cuales, además, deberían cubrir de forma generalizada todo el abanico de posibles patologías, derivaciones y condiciones en las que se pueden generar. En nuestro caso, las señales utilizadas en este trabajo provienen de la base de datos del MIT [MM98]. Es por ello por lo que en este apartado se describen tanto las características que debe tener una base de datos de señales ECG así como las principales bases de datos utilizadas y sobre las que se han validado los desarrollos realizados.

2.6.1. Bases de datos ECG

En general, las características exigibles a las bases de datos electrocardiográficas [GAG⁺00] se pueden resumir en los siguiente puntos:

1. Deberán contener señales representativas para llevar a cabo pruebas de algoritmos analizadores de ECGs, por lo que es necesario disponer de un gran conjunto de señales reales.
2. Deberán contener señales raramente observadas pero clínicamente significativas. Aunque no es difícil obtener registros electrocardiográficos correspondientes a ciertas anomalías, a menudo aquellos más significa-

tivos son raramente registrados. Para el desarrollo de analizadores de señal electrocardiográfica es necesario tener en cuenta estos casos.

3. Deberán contener señales estándar, es decir, señales utilizadas por un amplio sector de la comunidad científica relacionada con el tratamiento de señal electrocardiográfica. No tiene sentido el realizar comparaciones entre algoritmos si éstos no han sido aplicados al mismo conjunto de datos, ya que los resultados obtenidos dependen en muchos de los casos del conjunto de datos empleado.
4. Deberán contener señales con anotaciones. Normalmente, cada complejo *QRS* ha sido manualmente anotado por dos o más cardiólogos trabajando de forma independiente. Estas anotaciones sirven de referencia para comparar los resultados producidos por cierto método automático de procesamiento, y los producidos según el estándar de los cardiólogos.
5. Deberán contener señales discretas accesibles a programas informáticos. De esta forma, es posible llevar a cabo un test completamente automático y reproducible, permitiendo comparar resultados frente a modificaciones en los algoritmos utilizados.
6. Los parámetros de las señales tales como f_s (frecuencia de muestreo), ganancia, amplitud, etc, deberán ser conocidos.

Estas bases de datos permitirán una evaluación de los métodos desarrollados reproducible, automática, cuantitativa y estándar. Sin embargo, habrá que tener en cuenta que la utilización de una sola base de datos puede dar lugar a errores al querer extrapolar los métodos testados en ella sobre señales reales, ya que éstos pueden haberse adaptado de una manera excesiva al conjunto de datos empleados en su desarrollo. Por esta razón se aplicarán todos los métodos desarrollados a señales provenientes de distintos conjuntos. En algunos casos, será admisible la utilización de señales sintéticas que realcen ciertos aspectos a estudiar o simplemente permitan controlar mejor las características de las muestras.

En cuanto a las bases de datos de señales electrocardiográficas disponibles en la actualidad, podemos citar las que siguen [GAG⁺00]:

1. *AHA DB*. The American Heart Association Database for Evaluation of Ventricular Arrhythmia Detectors (80 registros de 35 minutos cada uno).
2. *MIT/BIH DB*. The Massachusetts Institute of Technology - Beth Israel Hospital Arrhythmia Database (48 registros de 30 minutos cada uno).
3. *ESC DB*. The European Society of Cardiology ST-T Database (90 registros de 120 minutos cada uno).
4. *NST DB*. The Noise Stress Database (12 registros de 30 minutos cada uno).
5. *CUIDB*. The Creighton University Sustained Ventricular Arrhythmia Database (35 registros de 8 minutos cada uno).

Sin embargo, a través del MIT (Massachusetts Institute of Technology) se tiene acceso a registros del resto de bases de datos, con lo cual todas las señales a utilizar se pueden obtener de la misma fuente. Con el paso del tiempo se han ido añadiendo registros pertenecientes a otras patologías, y de mayor duración, llegando a incluir incluso registros de hasta 24 h. Así, utilizando como fuente la base de datos del MIT, se dispone de un conjunto de señales lo suficientemente amplio para llevar a cabo pruebas muy representativas, además de contar con un formato de datos abierto y común a todas ellas, lo que reduce enormemente el esfuerzo en el desarrollo de aplicaciones. Éstas van a ser las razones principales por lo que va a ser la fuente utilizada en los experimentos llevados a cabo para validar los métodos propuestos en capítulos posteriores.

2.6.2. La base de datos del MIT

Como ya se ha comentado anteriormente y para poder validar los algoritmos de procesamiento desarrollados, será necesaria su aplicación sobre un amplio conjunto de señales obtenidas a partir de situaciones reales. Para ello y en este caso, se han utilizado como fuente de señales electrocardiográficas las bases de datos ofrecidas en la web¹ de PhysioNet [GAG⁺00].

El *Research Resource for Complex Physiologic Signals*, al que pertenece Physionet,

¹www.physionet.org

es un proyecto de investigación iniciado en Boston por instituciones médicas tan relevantes como el *Beth Israel Hospital* (BIH) (llamada ahora *Beth Israel Deaconess Medical Center*), la *Harvard Medical School*, la *Boston University*, *McGill University* y el *Massachusetts Institute of Technology* (MIT) y fundado bajo los auspicios del *National Center for Research Resources* y el *National Institute of Health*. Los recursos para la investigación y el estudio sobre señales complejas de origen biológico y fisiológico (no sólo señales electrocardiográficas) ofrecidos por estas instituciones a través de la web anteriormente citada, se distribuyen públicamente a través de tres componentes fuertemente interrelacionados:

[PhysioNet.] Se trata de un forum *on-line* ideado para la puesta en común y el intercambio de todo tipo de señales biomédicas y del software (open-source) adecuados para el análisis de las señales biomédicas, facilitando de esta forma la cooperación para el estudio de este tipo de datos y en la evaluación y prueba de nuevos algoritmos.

[PhysioToolkit.] Donde se presenta una recopilación de software para el procesamiento y análisis de señales de origen fisiológico y para la detección de eventos fisiológicamente significativos. Hay que tener en cuenta que uno de los principales temas que unifica varias líneas de investigación y que supone un importante aumento del software disponible en PhysioToolkit es el de la extracción de la información oculta que reside en las señales biomédicas y que puede resultar de gran utilidad en el diagnóstico precoz de múltiples enfermedades.

[PhysioBank.] Constituido por un archivo creciente de registros digitales correspondientes a señales de origen fisiológico que se ofrece libremente a través de la web a la comunidad científica internacional que investiga y trabaja sobre ellas (*Figura 2.6*). PhysioBank incluye en la actualidad una gran cantidad de bases de datos distintas con señales biomédicas que tratan de abarcar un amplio espectro de casos reales provenientes tanto de pacientes sanos como de pacientes con deficiencias en su salud (presentando casos por muerte súbita, fallos cardíacos, epilepsia, apnea, etc.). En la actualidad PhysioBank aglutina unas 30 bases de datos de acceso público gratuito. Dentro de las bases de datos de señales electrocardiográficas podemos encontrar las que se relacionan a continuación:



Figura 2.6: Aspecto de la web de PhysioBank

- *ANSI/AAMI EC13 Test Waveforms*. Base de datos cuyos registros se utilizan para probar y calibrar los diferentes dispositivos utilizados en la monitorización de las señales electrocardiográficas. Está compuesta por 10 registros de corta duración especificados por la *American National Standard*. Los registros son de dominio público e incluyen tanto señales sintéticas como reales. Cada uno de ellos contiene una señal electrocardiográfica obtenida mediante muestreo a 720 Hz con una resolución de 12 bits.
- *European ST-T Database*. Utilizada para evaluar todos aquellos algoritmos desarrollados para el análisis de los cambios ocurridos durante el segmento *ST* y la onda *T* de una señal electrocardiográfica. Está formada por 90 fragmentos representativos, perfectamente etiquetados (por dos cardiólogos diferentes) y seleccionados a partir de las señales extraídas de los electrocardiogramas registrados sobre 79 pacientes (70 hombres, 8 mujeres y un registro no identificado). En cada uno de los sujetos se ha diagnosticado un posible problema de isquemia de miocardio, además de utilizar otros criterios de selección sobre los fragmentos para

obtener una colección de disfunciones cardíacas lo suficientemente representativa (como, por ejemplo, el desplazamiento de la línea de base del segmento *ST*). La base de datos proporciona un total de 367 episodios de cambio de la línea de base del segmento *ST*, y 401 episodios de cambio de la onda *T*, con duraciones que oscilan entre los 30 segundos a varios minutos y con oscilaciones en la amplitud de la señal que van desde los $100 \mu\text{V}$ hasta más de 1 mV. Adicionalmente también se han etiquetado 11 episodios de desplazamiento axial que resultan en un cambio aparente del segmento *ST* y otros 10 episodios de desplazamiento axial que suponen cambios en la onda *T*. Cada uno de los registros presentados contiene dos señales de 2 horas de duración, muestreadas a 250 Hz con una resolución de 12 bits y un rango de entrada nominal de 20 mV.

- *Long-Term ST Database*. Se trata de una base de datos [JMT⁺96] [JTM⁺03] concebida para recoger toda la variabilidad que presenta el segmento *ST* en los electrocardiogramas reales, incluyendo episodios de isquemia, episodios no isquémicos y relacionados con el movimiento axial, episodios de lentitud en el desplazamiento del segmento *ST* y cualquier variante de la mezcla de estos tres. Está formada por 86 registros de larga duración (de 21 a 24 horas, razón por la cual aparecen fragmentados) conteniendo 2 ó 3 señales electrocardiográficas que han sido etiquetadas latido a latido. Los registros han sido extraídos a partir de 80 pacientes. Precisamente debido a la larga duración de los registros únicamente son accesibles desde Physionet 43 de ellos.
- *MIT-BIH Arrhythmia Database*. En esta base de datos se recogen gran cantidad de episodios de arritmia cardíaca en 48 extractos de media hora de duración correspondientes a los ECG ambulatorios de 47 pacientes, 23 registros seleccionados de forma aleatoria a partir de 4000 ECG ambulatorios de 24 horas de duración, y otros 25 registros elegidos para recoger también aquellas arritmias menos frecuentes pero clínicamente significativas. Los registros (correspondientemente etiquetados) han sido muestreados a 360 Hz con una resolución de 11 bits y con un rango de tensión de 10 mV. La mitad de los registros son accesibles a través de Physionet.
- *MIT-BIH Noise Stress Test Database*. Con 12 registros ECG y 3 registros que

contienen el ruido típico que aparece en los ECG ambulatorios [MMM84]. Todos ellos tienen una duración de media hora. Los 3 registros *ruidosos* han sido obtenidos ensamblando aquellos intervalos de registros que contenían tanto variaciones en la línea de base, como interferencias de carácter muscular o debidas al movimiento de los electrodos.

- *BIDMC Congestive Heart Failure Database*. Incluye ECGs de larga duración correspondientes a 15 pacientes que presentan fallos cardíacos congestivos de tipo severo [TLJ⁺01] [MPH⁺02]. Los registros tienen una duración de 20 horas, conteniendo 2 señales muestreadas a 250 Hz con una resolución de 12 bits y un rango de ± 10 mV.
- *Post-Ictal Heart Rate Oscillations in Partial Epilepsy*. Base de datos con los registros utilizados para el estudio referido en [AAKH⁺99].
- *QT Database*. La base de datos QT [LMGM97] fue diseñada a partir de la base de datos del MIT con el objetivo de evaluar algoritmos de detección de ondas en el ECG. Esta base de datos consiste en 105 fragmentos de 15 minutos de duración, provenientes de registros *Holter* de dos canales, seleccionados por varios expertos con el objetivo de incluir una amplia variedad de complejos *P*, *QRS*, *T* y *U* que aparecen adecuadamente etiquetados e identificados (de 30 a 50 latidos etiquetados de este modo en cada registro). De esta forma, las técnicas que se apliquen sobre estas señales estarán trabajando dentro de la variabilidad del mundo real. Los límites de cada onda dentro de estos registros han sido determinados manualmente por expertos. Los registros, que pertenecen a bases de datos ya existentes (incluyendo la base de datos de arritmias del MIT, la base de datos ST-T de la sociedad europea de cardiología y algunas otras y cuya distribución se muestra en el *Cuadro 2.4*) han sido elegidos para representar los extremos de patologías cardíacas.

MIT-BIH Arrhyth.	MIT-DB ST DB	MIT-BIH Sup. Vent.	MIT-BIH Long Term	ESC STT	MIT-BIH NSR DB	Sudden Death
15	6	13	4	33	10	24

Cuadro 2.4: Composición de la base de datos QT

- *Creighton University Ventricular Tachyarrhythmia Database*. Incluye 35 registros ECG de aproximadamente 8 minutos de duración, correspondientes a pacientes que han experimentado episodios de taquicardia ventricular, flutter ventricular y fibrilación ventricular. Todos los registros han sido muestreados a 250 Hz con 12 bits de resolución y un rango dinámico de 10 V (10 mV en las señales antes de ser amplificadas). Todos los registros son de acceso libre a través de Physionet.
- *Intracardiac Atrial Fibrillation Database*. Esta base de datos está formada por los registros endocardiales provenientes de la aurícula derecha de 8 pacientes en episodios de fibrilación o palpitación (flutter) auricular. Los registros han sido obtenidos a través de un catéter multipolar situado en cuatro regiones separadas del corazón. La frecuencia de muestreo es de 1 KHz. Todas las señales de la base de datos son accesibles a través de la web.
- *MIT-BIH Atrial Fibrillation Database*. Incluye 25 registros ECG de larga duración correspondientes a pacientes que presentan episodios de fibrilación auricular. Los registros de la base de datos tienen una duración aproximada de 10 horas y presentan 2 señales electrocardiográficas muestreadas a 250 Hz con una resolución de 12 bits y un rango de ± 10 mV [MM83]. Todas las señales están etiquetadas.
- *MIT-BIH ECG Compression Test Database*. Que contiene 168 registros ECG de corta duración (20.48 s cada uno de ellos) seleccionados para proporcionar un conjunto de señales test que abarca toda una amplia gama de posibilidades que aparecen en los casos reales y que deben ser resueltas adecuadamente por los diferentes métodos de compresión de señales electrocardiográficas, en particular aquellas técnicas de compresión con pérdidas [MMG88].
- *MIT-BIH Long-Term Database*. Esta base de datos proporciona 7 registros ECG de larga duración (entre 14 y 22 horas de duración cada uno de ellos) que han sido etiquetados manualmente.
- *MIT-BIH Malignant Ventricular Arrhythmia Database*. Con 22 registros de media hora de duración pertenecientes a pacientes que han presenta-

do episodios de taquicardia ventricular sostenida, flutter ventricular y fibrilación ventricular [GAMM85] [Gre86].

- *MIT-BIH Normal Sinus Rhythm Database*. Proporciona 18 registros ECG de larga duración pertenecientes a 18 pacientes (5 hombres y 13 mujeres) que no presentan ningún tipo de arritmia significativa.
- *MIT-BIH ST Change Database*. En la que se incluyen 28 registros ECG de diferentes duraciones, la mayoría de los cuales han sido obtenidos mientras los pacientes realizaban pruebas de esfuerzo para poder reflejar claramente las variaciones del segmento *ST* durante los cambios del ritmo cardíaco [Alb83].
- *MIT-BIH Supraventricular Arrhythmia Database*. En la que se proporcionan 78 registros de media hora de duración que han sido especialmente seleccionados para complementar los ejemplos de arritmia supraventricular proporcionados en la *MIT-BIH Arrhythmia Database* [Gre90].
- *Paroxysmal Atrial Fibrillation Prediction Challenge Database*. Esta base de datos fue creada con el objetivo de utilizar sus registros para desarrollar métodos automatizados para la detección de la fibrilación auricular de tipo paroximal. Contiene registros ECG de dos canales muestreados a 128 Hz con una resolución de 16 bits por muestra.

2.7. Características y composición de las fuentes

Para evaluar las técnicas propuestas en capítulos posteriores sobre un conjunto de señales reales se han seleccionado una serie de ECGs pertenecientes a las bases de datos del MIT [GAG⁺00], todas ellas accesibles a través de la aplicación PhysioBank. La mayoría de las señales han sido muestreadas con una frecuencia de 360 Hz, una resolución de 12 bits y un rango de tensión de ± 5 mV donde un valor de muestra de 1024 equivale a una tensión de 0 V (*Cuadros 2.5 y 2.6*). En total se han utilizado 49 señales conteniendo unos 49318 latidos con un volumen de información procesada de 15.5 MB.

Por otra parte, en los *Cuadros 2.7 y 2.8* se describe la composición en latidos de las ECGs utilizadas. Los latidos han sido segmentados manualmente

y etiquetados por personal cualificado. En nuestro caso se ha prescindido del valor médico de las etiquetas, identificando cada tipo de latido del ECG por un determinado símbolo, donde los etiquetados con el identificador 99 corresponden a latidos corruptos cuya forma de onda no se corresponde con ninguna de las clases esperadas y que definiremos como outliers. Estas etiquetas sólo se aplican al ámbito del ECG analizado, ya que no mantienen el mismo significado sobre registros diferentes.

Registro	Número de latidos	Número de muestras	Bits por muestra	Frecuencia (Hz)	Tamaño (kB)
10	356	72058	12	250	105
11	355	72058	12	250	105
12	1017	224584	12	250	328
13	1035	224585	12	250	328
101	960	170901	11	360	250
102	554	163160	11	360	239
103	556	163161	11	360	239
104	541	159875	11	250	234
105	540	159876	11	360	234
106	661	224288	11	360	328
107	661	224286	11	360	328
108	2267	649624	12	250	951
109	2272	649627	12	250	951
110	356	72058	12	250	105
111	355	72058	12	250	105
112	1017	224584	12	250	328
113	1035	224585	12	250	328
200	2685	252363	11	360	369
201	660	215697	11	360	315
202	659	215699	11	360	315
203	956	85231	11	360	124
204	703	85144	12	250	124
205	1201	224831	11	360	329
206	1202	224829	12	250	329
207	1026	224643	11	360	329
208	1026	224644	11	360	329
209	1424	219036	11	360	320
210	1407	219033	11	360	320

Cuadro 2.5: Características principales de los registros electrocardiográficos (i).

Registro	Número de latidos	Número de muestras	Bits por muestra	Frecuencia (Hz)	Tamaño (kB)
211	1136	224717	12	250	329
212	1124	224718	11	360	329
213	1639	224763	11	360	329
214	1635	224764	11	360	329
300	1307	224730	12	250	329
301	1303	224590	12	250	328
1000	956	170901	12	250	250
1001	960	170901	12	250	250
1002	554	163160	12	250	239
1003	556	163161	12	250	239
1004	541	159875	12	250	234
1005	540	159876	12	250	234
1006	661	224288	12	250	328
1007	661	224286	12	250	328
1008	2267	649624	12	250	951
1009	2272	649627	12	250	951
100	956	170901	11	250	250
1010	356	72058	12	250	105
1011	355	72058	12	250	105
1012	1017	224584	12	250	328
1013	1035	224585	12	250	328

Cuadro 2.6: Características principales de los registros electrocardiográficos (ii).

Registro	a	b	c	d	e	f	99
10	249	56	42	6	0	0	3
11	251	64	40	0	0	0	0
12	951	43	23	0	0	0	0
13	957	56	17	0	0	0	5
100	934	7	9	0	0	0	6
101	936	21	0	0	0	0	3
102	554	0	0	0	0	0	0
103	254	0	0	0	0	0	2
104	538	1	0	0	0	0	2
105	538	1	0	0	0	0	1
106	660	0	0	0	0	0	1
107	660	1	0	0	0	0	0
108	2265	1	0	0	0	0	1
109	2270	1	0	0	0	0	1

Cuadro 2.7: Composición en latidos de las fuentes electrocardiográficas utilizadas (i).

Registro	a	b	c	d	e	f	99
110	249	56	42	6	0	0	3
111	251	64	40	0	0	0	0
112	951	43	23	0	0	0	0
113	957	56	17	0	0	0	5
200	1632	212	4	19	2	688	128
201	519	139	0	0	0	0	2
202	519	140	0	0	0	0	0
203	175	256	51	369	1	1	103
204	515	175	6	0	0	0	7
205	1141	59	0	0	0	0	1
206	1143	59	0	0	0	0	0
207	999	27	0	0	0	0	0
208	998	27	0	0	0	0	1
209	1401	2	4	0	0	0	17
210	1403	4	0	0	0	0	0
211	1109	6	1	0	0	0	20
212	1116	7	0	0	0	0	1
213	11	1415	10	199	0	0	4
214	1442	184	1	0	0	0	8
300	1012	215	18	61	0	0	1
301	1069	204	21	2	1	0	6
1000	934	7	9	0	0	0	6
1001	936	21	0	0	0	0	3
1002	554	0	0	0	0	0	0
1003	554	0	0	0	0	0	2
1004	538	1	0	0	0	0	2
1005	538	1	0	0	0	0	1
1006	560	0	0	0	0	0	1
1007	660	1	0	0	0	0	0
1008	2265	1	0	0	0	0	1
1009	2270	1	0	0	0	0	1
1010	249	56	42	6	0	0	3
1011	251	64	40	0	0	0	0
1012	951	43	23	0	0	0	0
1013	957	56	17	0	0	0	5

Cuadro 2.8: Composición en latidos de las fuentes electrocardiográficas utilizadas (ii).

Capítulo 3

Estado del Arte

***Resumen.** La elevada cantidad de aplicaciones existentes para el procesamiento de señales de origen fisiológico es principalmente debida a la complejidad en la extracción de reglas y características concretas para la implementación de algoritmos que reflejen de forma inequívoca el conocimiento médico que se deriva de la interpretación de las señales biológicas tratadas. En este capítulo se abordará la revisión bibliográfica de los métodos y técnicas utilizadas en las etapas de preprocesamiento (adquisición de datos, eliminación de interferencias, aproximación poligonal y segmentación de la señal), así como todas aquellas fuentes relativas al proceso de selección y extracción de características y técnicas de reconocimiento de patrones, situando, de esta manera, el contexto adecuado a todas las aportaciones realizadas. El capítulo termina con un apartado de conclusiones.*

Contenidos

3.1. Técnicas de preprocesamiento	39
3.1.1. Ruido	39
3.1.2. Interferencia de la red	48
3.1.3. Variaciones de la línea base	51
3.2. Extracción de características	56
3.3. Segmentación	57
3.3.1. Detección de las ondas <i>P</i> y <i>T</i>	57
3.3.2. Detección del complejo <i>QRS</i>	58
3.4. Reducción de características	64
3.4.1. Análisis de componentes principales	64
3.4.2. Modelos ocultos de Markov	67
3.5. Reconocimiento de patrones	92
3.5.1. Métodos supervisados	93
3.5.2. Métodos no supervisados	94
3.6. Conclusiones	105

3.1. Técnicas de preprocesamiento

Una vez adquirida la señal deberemos prepararla adaptando sus características para su posterior tratamiento informático. Esto es debido a que la señal obtenida tras el proceso de adquisición (*Ecuación 3.1*) presenta una serie de elementos ajenos a la propia señal (ruido, interferencia de la red de alimentación o variaciones de la línea base). Todos estos elementos que enmascaran la información, deberán ser aislados y minimizados para que, en las siguientes fases de análisis, los resultados sean independientes de todos estos factores.

$$x[n] = ECG[n] + r[n] + b[n] \quad (3.1)$$

donde $x[n]$ representa la señal discreta adquirida, $ECG[n]$ es la señal electrocardiográfica real producida por la actividad cardíaca que se desea registrar, $r[n]$ es el ruido en general (incluyendo algunos artefactos y la interferencia de la red eléctrica), y $b[n]$ son las variaciones de la línea base. En todos los casos n representa el índice temporal discreto.

El objetivo de los algoritmos de preprocesamiento [Cue99] es el de reducir al máximo $r[n]$ y $b[n]$ para que la señal obtenida $x[n]$ sea lo más parecida a la señal real $ECG[n]$. Para ello se realiza una estimación de dichas señales $\hat{r}[n]$ y $\hat{b}[n]$, de manera que, al restarlas a $x[n]$ en la *Ecuación 3.1*, se compense su efecto. En los siguientes apartados se describen tanto $r[n]$ como $b[n]$ así como los métodos más importantes existentes para minimizar sus efectos y conseguir que $x[n] \rightarrow ECG[n]$.

3.1.1. Ruido

El ruido es un elemento que se encuentra presente en mayor o menor medida en todas las funciones reales. En el caso que nos ocupa, el ruido que aparece en la señal electrocardiográfica puede ser debido a múltiples factores: desde el ruido debido al movimiento existente entre el electrodo y la piel, el ruido generado por los propios aparatos electrónicos utilizados en la fase de adquisición (ruido térmico), hasta el ruido debido a la generación de otras señales de tipo eléctrico por parte de los músculos del cuerpo humano (señales electromiográficas), etc. Todo ello supone que, sobre la señal electrocardiográfica original se

superponga otra señal de cierto nivel que puede llegar a enmascarar rasgos significativos de la misma, tal y como se muestra en la *Figura 3.1*.

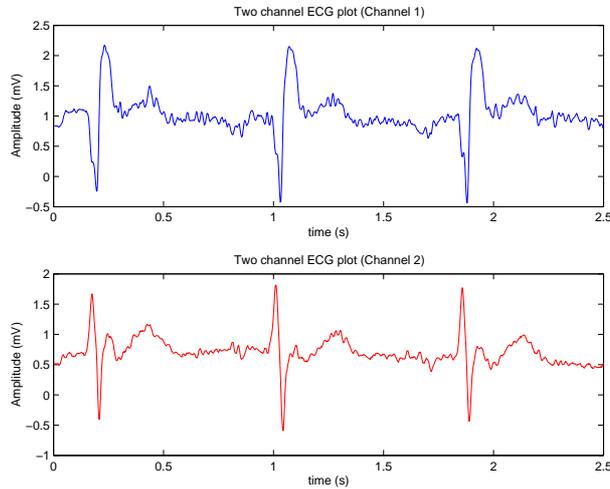


Figura 3.1: Señal ECG de doble registro donde aparece ruido superpuesto. La fuente utilizada procede de la base de datos del MIT-BIH *Noise Stress Test*.

En los siguientes apartados se exponen brevemente los principales métodos de reducción del ruido:

Métodos basados en técnicas clásicas de filtrado.

Uno de los métodos más usuales de reducción del ruido presente en una señal lo constituye el uso de filtros paso bajo [AT85]. Se trata de un método genérico de reducción del ruido de una señal cualquiera [OS89], teniendo en cuenta que sus componentes frecuenciales están por encima de las de la señal. Estos filtros se diseñan utilizando las técnicas clásicas de diseño de filtros FIR (respuesta al impulso finita) e IIR (respuesta al impulso infinita). Debido a su escasa selectividad (sobre todo si se diseñan con pocos coeficientes) no se suelen utilizar en el caso de señales biomédicas. Las técnicas de filtrado lineal se definen en el dominio temporal mediante la operación de convolución (*Ecuación 3.2*)

$$y[n] = \sum_{k=-\infty}^{\infty} h[n-k]x[k] = x[n] * h[n] \quad (3.2)$$

donde $x[n]$ representa la señal de entrada (en este caso las muestras de la señal electrocardiográfica capturadas desde el Holter); $h[n]$ sería el filtro diseñado para la eliminación del ruido; e $y[n]$ representa la señal de salida una vez filtrada (Figura 3.2).

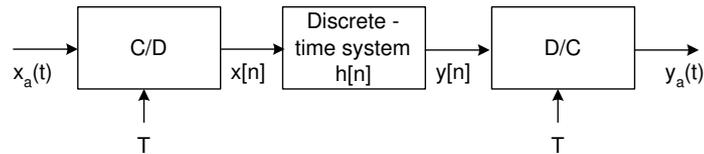


Figura 3.2: Sistema básico para el filtrado discreto de señales continuas donde la frecuencia de muestreo $f_s = \frac{1}{T_s}$.

Para un sistema discreto y si la señal de entrada está limitada en banda (tal y como lo están las señales electrocardiográficas) y con una frecuencia de muestreo f_s lo suficientemente elevada para evitar el aliasing, entonces el sistema entero mostrado en la Figura 3.2 se comporta como un sistema lineal continuo invariante en el tiempo y con la siguiente respuesta frecuencial:

$$H(j\Omega) = \begin{cases} H(e^{j\Omega T}) & |\Omega| < \pi/T_s \\ 0 & |\Omega| > \pi/T_s \end{cases}$$

donde Ω responde a la frecuencia en el dominio continuo equivalente a $2\pi f$. El paso de las especificaciones del filtro continuo en el tiempo a las del mismo filtro en un dominio discreto se efectúa a través de la relación $\omega = \Omega T_s$, con lo que:

$$H(e^{j\omega}) = H(j\frac{\omega}{T}), \quad |\omega| < \pi \quad (3.3)$$

Por ejemplo, en [AT85] se describen dos filtros paso bajo para eliminar el ruido de alta frecuencia, uno de ellos de primer orden (Ecuaciones 3.4 y 3.5), y

el otro de segundo orden, en el que se mejoran la presencia de lóbulos secundarios y la poca selectividad del primero de los filtros presentados (Ecuaciones 3.6 y 3.7).

$$H(z) = \frac{1 - z^{-m}}{1 - z^{-1}} \quad (3.4)$$

$$y[n] = y[n - 1] + x[n] - x[n - m] \quad (3.5)$$

$$H(z) = \frac{(1 - z^{-4})^2}{(1 - z^{-1})^2} \quad (3.6)$$

$$y[n] = 2y[n - 1] - y[n - 2] + x[n] - 2x[n - 4] + x[n - 8] \quad (3.7)$$

En la *Figura 3.3* se observa el espectro de una señal electrocardiográfica, donde se puede apreciar el contenido frecuencial máximo de este tipo de señales (normalmente por debajo de los 50 Hz aproximadamente).

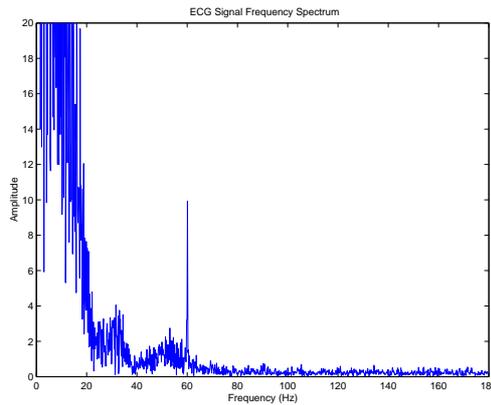


Figura 3.3: Espectro de una señal electrocardiográfica real. La fuente utilizada procede del MIT-BIH *Arrhythmia Database*. Además en la figura se puede apreciar la interferencia de la red, que está localizada en los 60 Hz.

Para el diseño de un filtro paso bajo que elimine el ruido a partir de cierta frecuencia, será necesario el enventanado de la respuesta impulsional infinita del filtro paso bajo ideal, por lo que se utilizan como interpoladores filtros FIR

diseñados por el método de las ventanas. Algunas de las características de los filtros de ventana más frecuentemente utilizados son las que se muestran en el *Cuadro 3.1* (donde M es el ancho de la ventana en número de muestras).

Cuadro 3.1: Comparativa entre algunos de los filtros de ventana.

Filtro	Amplitud de pico del lóbulo secundario	Ancho aprox. lóbulo principal	Aproximación del error de pico (dB)	Ventana Kaiser equiv.	Ancho de transición de la Kaiser equivalente
Rectangular	-13	$4\pi/(M+1)$	-21	0	$1,81\pi/M$
Bartlett	-25	$8\pi/M$	-25	1.33	$2,37\pi/M$
Hanning	-31	$8\pi/M$	-44	3.86	$5,01\pi/M$
Hamming	-41	$8\pi/M$	-53	4.86	$6,27\pi/M$
Blackman	-57	$12\pi/M$	-74	7.04	$9,19\pi/M$

Métodos basados en filtros de media móvil.

Otra forma de minimizar el ruido es usando filtros de media móvil [OS89] [Oka79]. En este caso, se trata de un método de filtrado muy simple aunque con baja selectividad, además de enmascarar detalles en la alta frecuencia de la señal. Estos filtros tienen una expresión genérica:

$$y[n] = \frac{1}{a+b+1} \sum_{k=n-a}^{k=n+b} x[k] \quad (3.8)$$

de manera que la muestra situada en el centro de un intervalo de anchura k , resulta de la ponderación de las muestras vecinas a la muestra original. Dependiendo de autores, se utiliza este filtro con unos valores de ventana de a y b distintos, además de aplicar distintos pesos a las muestras. Por ejemplo, en [Oka79], el filtro de media móvil que se propone es:

$$y[n] = \frac{x[n-1] + 2x[n] + x[n+1]}{4} \quad (3.9)$$

Métodos basados en el promedio de latidos.

Aprovechando tanto la semejanza entre latidos, como la semiperiodicidad y la duración (en número de latidos, para asegurar que se cuenta con un número suficiente de períodos) de este tipo de señales, se puede realizar un promedio de varios latidos semejantes para reducir el ruido. Para ello será necesario primeramente el segmentar la señal en sus latidos componentes [Oka79], así como aplicar algún tipo de alineamiento temporal [SSL96] para poder sumar latidos de distinta duración.

El promediado temporal se utiliza para mejorar la relación señal a ruido de señales electrocardiográficas de alta resolución. Para ello se hace uso de su característica repetitiva y de la no correlación entre el ruido y la señal. El uso del promediado requiere un punto de sincronismo como referencia, ya que el desalineamiento de las señales a promediar produce un efecto indeseable de filtrado paso bajo. En efecto, supongamos que tenemos una señal ideal $s(t)$ contaminada por una señal ruidosa $r(t)$, que se encuentra desalineada respecto del punto de alineamiento de la señal ideal. Si el desalineamiento del ruido τ_i es pequeño comparado con las variaciones de la señal, se puede hacer una aproximación de Taylor de la señal desplazada $r(t - \tau_i) = r(t) - r'(t) \cdot \tau_i$. Calculando ahora el cuadrado de la densidad de potencia (dp) como:

$$dp^2(t) = E[(s(t) + r(t - \tau_i) - E[s(t) + r(t - \tau_i)])^2] = r'^2 E[\tau_i^2] = r'^2 \sigma^2 \quad (3.10)$$

donde σ^2 es la desviación del desalineamiento. De aquí se deduce que la potencia de la señal desviación no depende sólo del desalineamiento del ruido, sino también de las características frecuenciales de éste. Cuanto mayores son las componentes frecuenciales, mayor es la potencia de desviación. A partir de aquí se proponen los métodos de Woody (sobre ruido blanco) y el de las integrales normalizadas (para la interferencia de red a 50 Hz) como los mejores para localizar el punto de sincronismo y resolver el problema del alineamiento de los latidos en registros ECG.

Métodos basados en la aproximación mediante funciones.

Otras técnicas basadas en el reconocimientos de formas utilizan ciertas aproximaciones de la señal como método para minimizar el ruido. Entre estas técnicas encontramos la del ajuste de una curva mediante segmentos [KJ96] (Figura 3.4) y el ajuste de curvas mediante funciones ortogonales y polinomios [OGJL97] [SBNP81], las cuales se utilizan igualmente como algoritmos de compresión de datos.

En [OGJL97] se describe la representación de esta señal mediante expansiones ortogonales de un reducido número de coeficientes utilizando el producto escalar y estimación adaptativa mediante el algoritmo LMS (mínimos cuadrados), donde el método elegido depende de la relación señal a ruido inicial.

Además, en el *Capítulo 4* se propone un método de aproximación poligonal basado en la diferencia entre ordenadas entre la señal original y su aproximación que, por su simplicidad resulta muy eficiente para la reducción del tiempo de procesamiento.

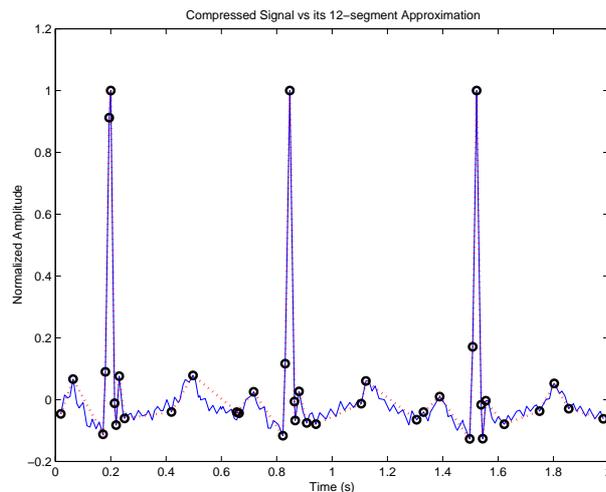


Figura 3.4: Ejemplo de la eliminación del ruido mediante la aproximación poligonal de un latido: la aproximación lineal tiene el mismo efecto que un filtrado paso bajo, minimizando el efecto del ruido.

Métodos basados en la transformada *wavelet*.

La transformada *wavelet* es una herramienta relativamente reciente en múltiples aplicaciones de procesamiento de señales, entre las que se encuentra la reducción del ruido. Para llevar a cabo esta reducción, prácticamente todos los métodos se basan en los siguientes pasos [Don96]:

1. Cálculo de la transformada *wavelet* de una señal hasta el nivel deseado.
2. Aplicación de un umbral y una función de umbralizado a los coeficientes del detalle, para eliminar en principio aquellos componentes que representan mayoritariamente el ruido (se consideran aquellos que no superan el valor dado por una función de umbralizado).
3. Cálculo de la transformada inversa para recuperar la señal, una vez eliminadas las componentes de ruido.

Uno de los trabajos más representativos de esta técnica es el presentado en [DJ95], en el que se basan la mayoría de los trabajos restantes. En éste se establecen los principios básicos para aplicar esta herramienta a la reducción del ruido, que son:

- La señal resultante del proceso debe ser al menos tan suave como la señal original.
- Se minimice el error cuadrático medio.

A partir de este trabajo se derivan otros muchos, donde se estudian variaciones del método anterior para conseguir mejorar la SNR.

Por ejemplo, en [IM98], se describe un método iterativo basado en la transformada *wavelet* (WT), añadiendo técnicas vectoriales de proyección sobre conjuntos y así evitar el solapamiento que existe entre los filtros de la transformación, lo cual provoca que la umbralización no sea la ideal.

En [KW98] se describen algunas modificaciones a la umbralización para aplicar el método en algunos casos en el que el ruido no corresponde a ruido blanco gaussiano, tal como se supone en la mayoría de ocasiones.

En [KB98] se utiliza un método basado en la correlación de tercer orden para

seleccionar el umbral, con lo cual se consigue que la SNR obtenida sea más constante respecto al umbral escogido, aunque en algunos casos los resultados obtenidos son peores que con el método clásico.

En [ea97] también aplica una variación a la selección del umbral. En este caso se utiliza la validación cruzada de splines para calcular este umbral. Como principal desventaja de esta variante destaca el coste computacional de los cálculos.

En [CW99] se utiliza un método iterativo basado en la selección de la mejor descomposición de una señal utilizando la transformada *wavelet*, a partir de la medida de la entropía obtenida variando la familia de *wavelets*.

En nuestro caso, tomaremos como punto de partida los resultados obtenidos en [Cue01] donde se utiliza el par DWT-IDWT (WT discreta y su función inversa) (Ecuaciones 3.12 y 3.14) para la reducción del ruido en señales biomédicas. En este caso, la expresión genérica de la señal de entrada es:

$$y[n] = x[n] + \sigma r[n] \quad (3.11)$$

donde $x[n]$ corresponde a la señal ECG sin ruido, $r[n]$ es el ruido, que en este trabajo se considerará como ruido blanco gaussiano de media nula y varianza unidad, lo cual supone una aproximación bastante cercana a la realidad [TWT84], $N(\mu, \sigma^2) = N(0, 1)$, y finalmente σ corresponde al nivel de este ruido. El proceso genérico para reducir el ruido de una señal utilizando la WT (Ecuación 3.12) es tomar los coeficientes de la aproximación *wavelet*, aplicar un umbral, y calcular la IWT (Ecuación 3.14) para obtener la reconstrucción de la señal sin ruido (se intentan identificar los componentes que contienen el ruido reconstruyendo la señal sin dichos coeficientes) [TWT84]

En el caso discreto y sea $y[n]$ la señal a analizar (Ecuación 3.11), su transformada *wavelet* viene dada por:

$$C[j, k] = \sum_{n \in \mathbb{Z}} y[n] \Psi_{j,k}[n] \quad (3.12)$$

donde $\Psi_{j,k}$ es una *wavelet* discreta definida como:

$$\Psi_{j,k}[n] = 2^{-\frac{j}{2}} \Psi[2^{-j}n - k] \quad (3.13)$$

La transformada inversa se define de forma similar como:

$$y[n] = \sum_{j \in Z} \sum_{k \in Z} C[j, k] \Psi_{j, k}[n] \quad (3.14)$$

Las condiciones exigibles a este proceso de reducción del ruido son que se minimice el error cuadrático medio (o lo que es lo mismo, que maximice la SNR resultante), y que la señal obtenida sea al menos tan suave como la señal inicial [Don96], para evitar que, aunque se cumpla la primera condición, aparezcan picos en la señal resultante, que se pueden confundir con información presente en la señal, o puede dar lugar a que se eliminen puntos característicos del ECG.

Existen algunos métodos propuestos [IM98], [KB98], [NCE⁺00], [Bez96] para la reducción del ruido basándose en los requisitos mencionados. La diferencia entre ellos es la forma de calcular el umbral, y la manera de aplicarlo a los coeficientes del detalle. En cuanto al cálculo del umbral, la mayoría se basan en modificaciones del propuesto en [Don96], que viene dado por la expresión $\delta = \sqrt{2 \log(N)} \hat{\sigma}$, siendo N el número de coeficientes del detalle considerado, y $\hat{\sigma}$ una estimación del nivel de ruido obtenida a partir de ciertos cálculos. El problema de estos métodos es que son demasiado conservadores y los resultados obtenidos en señales con alto nivel de ruido no son muy satisfactorios. Una vez conocido este valor de $\hat{\sigma}$, ya se puede calcular el umbral δ . La forma de aplicar este umbral a los coeficientes puede ser lineal o no lineal. En general, esta última ofrece mejores resultados, pudiendo a su vez dividirse en *soft-thresholding* o *hard-thresholding* [Cue01].

3.1.2. Interferencia de la red

Otro efecto que aparece con relativa frecuencia en señales electrocardiográficas es la superposición de una interferencia debida a la señal de la red (*Figura 3.5*).

Para reducir el efecto de este tipo de interferencia existen una serie de técnicas [OS89] basadas principalmente en el diseño de distintos tipos de filtros (por ejemplo, se puede utilizar un filtro elimina-banda centrado en la frecuencia de la interferencia de la red y con un ancho de banda lo suficientemente

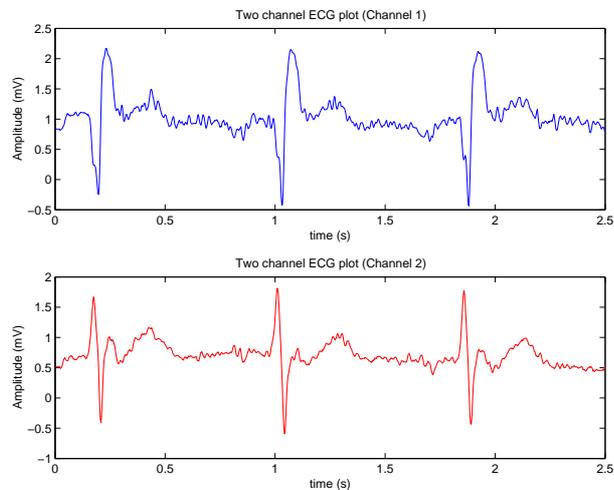


Figura 3.5: Señal ECG en la que aparece superpuesta la interferencia de la red (normalmente en Europa a unos 50Hz aunque ésta puede fluctuar).

estrecho como para que no perjudique al resto de componentes de la señal). En efecto, en la *Figura 3.3* se puede apreciar como, al representar en el dominio de la frecuencia la señal ECG original, aparece una interferencia claramente localizada en este caso en los 60 Hz (frecuencia de la red de alimentación en los E.E.U.U.).

En [PT95] se describe un método para reducir la interferencia de red en el dominio del tiempo utilizando un filtro *Notch* con los parámetros adecuados, aunque presentando los siguientes inconvenientes:

- No tiene en cuenta las fluctuaciones de la frecuencia interferente (ya que la frecuencia de la red no está localizada exactamente sino que va oscilando con el tiempo). Para evitar este efecto se propone en [FB94] un método adaptativo para la obtención de filtros de distribución gaussiana que se ajusten en todo momento a la frecuencia cambiante de la interferencia.
- Al aplicar el filtro propuesto aparece un transitorio al inicio de la señal que añade ruido a la misma, aunque para evitarlo también se propone en [PT95] un algoritmo adecuado.

El método utilizado para diseñar un filtro que se adapte a las variaciones de la frecuencia de la red, es el de emplazamiento de polos y ceros en el círculo unidad. El objetivo es ir cambiando los coeficientes del filtro para que se cumpla un criterio de optimización. La expresión genérica de este filtro es:

$$y[n] = \sum_{i=0}^N a_i x[n-i] - \sum_{j=1}^M b_j y[n-j] \quad (3.15)$$

con los coeficientes variando según:

$$\begin{aligned} a_i[n+1] &= a_i[n] + \mu \cdot e[n]x[n-i] \\ b_j[n+1] &= b_j[n] + \eta \cdot e[n]y[n-j] \end{aligned} \quad (3.16)$$

donde μ y η son parámetros de ajuste y $e[n]$ es la función escalón.

Finalmente en [RV03] se propone la construcción de una estructura de celdas con filtros pasa-todo utilizando filtros digitales IIR *Notch* de segundo orden. Este tipo de estructuras se caracterizan por su excelente sensibilidad a la frecuencia de corte y su alta estabilidad, consiguiendo un filtro con una transición muy abrupta con lo que se elimina perfectamente la interferencia de la red para la que ha sido diseñado. Además, utilizando este tipo de estructuras, también se consiguen eliminar las bajas frecuencias que ensucian la señal ECG al introducir una línea de base. La función de transferencia utilizada para el filtro digital IIR de tipo *Notch* que elimina tanto la interferencia de la red como la línea de base es:

$$H(z) = \frac{1}{2}[1 + A(z)] \quad (3.17)$$

donde:

$$A(z) = \frac{a_1 + a_2 z^{-1} + z^{-2}}{1 + a_2 z^{-1} + a_1 z^{-2}} \quad (3.18)$$

La estructura de celdas implementada para la obtención de $A(z)$ (Ecuación 3.18) es la mostrada en la *Figura 3.6*.

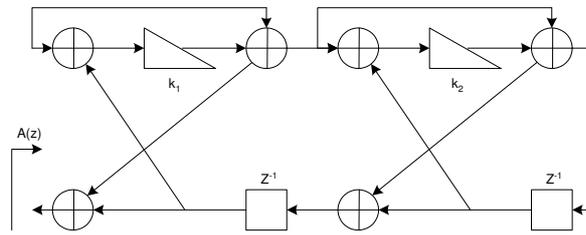


Figura 3.6: Estructura de filtros IIR Notch utilizados tanto para eliminar la interferencia de la red como la línea de base.

Y la función de transferencia correspondiente a esta estructura (Figura 3.6) resulta:

$$A(z) = \frac{k_2 + k_1(1 + k_2)z^{-1} + z^{-2}}{1 + k_1(1 + k_2)z^{-1} + k_2z^{-2}} \quad (3.19)$$

donde los coeficientes k_1 y k_2 referidos a la frecuencia digital del filtro (ω_0) y a su ancho de banda (Ω) se definen como:

$$k_1 = -\cos(\omega_0) \quad (3.20)$$

$$k_2 = \frac{1 - \tan(\Omega/2)}{1 + \tan(\Omega/2)} \quad (3.21)$$

3.1.3. Variaciones de la línea base

Las variaciones de la línea base se pueden producir debido a múltiples factores como, por ejemplo, al movimiento del paciente durante la adquisición del electrocardiograma, a la respiración y a cambios en la impedancia de los electrodos. Estas variaciones suponen una interferencia de baja frecuencia y de cierta amplitud que debe ser reducida (e idealmente eliminada) para no alterar el resultado del procesado posterior (Figura 3.7).

Estas componentes de baja frecuencia pueden inducir a error cuando se realiza una interpretación visual o análisis automático de la señal electrocardiográfica. El contenido frecuencial de estas variaciones normalmente está en

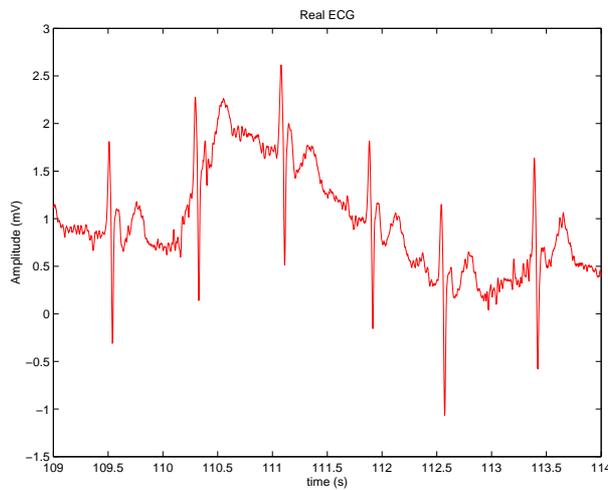


Figura 3.7: Señal ECG en la que aparece una interferencia debida a las variaciones de la línea base.

el rango de 0 a 0,5 Hz, aunque en las señales generadas para el test de esfuerzo el rango frecuencial suele ser mayor. Las técnicas más comunes que se pueden encontrar en la literatura especializada para reducir las variaciones de la línea base se pueden clasificar en los siguientes apartados:

Filtros clásicos.

Una primera solución podría ser la utilización de filtros paso banda de 0.5 Hz a 100 Hz, o de paso alto con una frecuencia de corte de 0.5 Hz, ya que estas interferencias normalmente tienen un contenido frecuencial por debajo de este valor. Sin embargo, esto no cumple las especificaciones de la American Heart Association (AHA) [MK97b], ya que se modifican componentes de baja frecuencia como el segmento *ST*. Además, los filtros lineales tienen un elevado coste computacional, y para conseguir una banda estrecha es necesario un elevado número de coeficientes [AS85]. Si se utilizan filtros IIR serán necesarios menos coeficientes, pero no tienen fase lineal, por lo que distorsionarán la señal.

Por ejemplo, en [AT85], se proponen un conjunto de filtros digitales a aplicar

en todas las etapas de preprocesado de la señal ECG. Entre ellos se propone un filtro para reducir las variaciones de la línea base. Este filtro es un filtro paso alto IIR, con una frecuencia de corte de 0.5 Hz.

Filtros adaptativos.

En [JLTC92] se describe un método para la reducción de la línea base utilizando como técnica el filtrado adaptativo.

Para ello se propone un filtrado adaptativo en cascada de dos pasos. El primero consiste en un filtrado adaptativo constante, y el segundo es un filtrado tomando como referencia los instantes de aparición de cada latido. En ambos se utiliza el algoritmo LMS.

El primer filtrado pretende eliminar la componente DC y las frecuencias menores que aquellas donde aparece el segmento *ST*. Después de esto y teniendo en cuenta el comportamiento repetitivo del ECG, se aplica un filtrado adaptativo con impulsos correlados, adecuado para la estimación de señales periódicas o de señales que se repiten por bloques en el tiempo. Este filtrado elimina el remanente de las variaciones en la línea base que no estén correladas con el *QRS* para así preservar las componentes del *QRS* que sí lo estén.

Aproximaciones mediante funciones.

Se trata de un método más exacto que los anteriores y que permite la eliminación de la línea base de mayor frecuencia, aunque a costa de un elevado coste computacional [AS85] [OIEC95] [CNE⁺00].

En [MK97b] se explica un método para la reducción de la línea base mediante aproximación de la señal con splines cúbicos. Está basado en la interpolación entre los niveles isoelectrónicos estimados en los intervalos *P – R*. De esta forma, su precisión dependerá en gran medida de la precisión en la selección de los puntos adecuados, la cual muchas veces no es la más conveniente por la dificultad inherente a la localización de éstos. Otro inconveniente es que el método pierde prestaciones cuando los puntos entre los cuales se interpola están relativamente distanciados.

La estimación de la línea base en el intervalo $[0, T1]$ se realiza mediante un polinomio de orden 3 de la forma descrita en la Ecuación 3.22.

$$y(t) = y'''(0)\frac{t^3}{6} + y''(0)\frac{t^2}{2} + y'(0)t + y(0) \quad (3.22)$$

Para obtener los correspondientes coeficientes de esta ecuación, se aplican una serie de restricciones en los puntos inicial y final de cada uno de los periodos estimados. Estas restricciones son:

$$y(0) = y_0 \quad (3.23)$$

$$y'(0) = y'_0 \quad (3.24)$$

donde $y(0)$ se obtiene de ciertas condiciones de inicialización. En el punto final se debe verificar que:

$$y(T_1) = y_1 \quad (3.25)$$

y por criterios de estabilidad:

$$y'(T_1) = \frac{y_2 - y_0}{T_2} \quad (3.26)$$

de donde:

$$y'(t) = y'''(0)t^2 + y''(0)t + y'(0) \quad (3.27)$$

con lo cual se llega a un sistema de dos ecuaciones con dos incógnitas para obtener la solución $y(t)$. Sin embargo, la solución del problema es mejor calcularla utilizando operaciones matriciales para ir encadenando los diferentes tramos. De esta forma, se llega a una expresión recursiva de la forma:

$$\begin{bmatrix} y(N+1) \\ y'(N+1) \\ y''(N+1) \\ y'''(N+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \frac{1}{2} & \frac{1}{6} \\ 0 & 1 & 1 & \frac{1}{2} \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} y(N) \\ y'(N) \\ y''(N) \\ y'''(N) \end{bmatrix} \quad (3.28)$$

Una vez se calcula $y(t)$, ésta se resta a la señal original para reducir las variaciones de la línea base.

En cuanto a la aproximación mediante polinomios de Chebyshev, en [OIEC95]

se describe un método similar al utilizado con la aproximación mediante Wavelets [CNE⁺00]. Para ello se basa en el hecho de que cualquier función continua puede ser aproximada por un conjunto de funciones polinómicas $\{\phi_k(t) \mid k = 0, 1, \dots, n\}$, donde el subíndice k representa el grado del polinomio. En este caso el objetivo final es aproximar la línea base $B(t)$ de la forma:

$$B(t) = c_0\phi_0(t) + c_1\phi_1(t) + c_2\phi_2(t) + \dots + c_n\phi_n(t) = \sum_{k=0}^n c_k\phi_k(t) \quad (3.29)$$

El conjunto de funciones escogido corresponde a los polinomios de Chebyshev definidos como:

$$\phi_k(t) = \cos(k \cos^{-1}(t)), \quad -1 \leq t \leq 1 \quad (3.30)$$

El objetivo de este método consiste en obtener los coeficientes y las funciones que ofrezcan la mejor aproximación de la señal. Una vez logrado esto, se calcula la curva resultante y se resta de la señal electrocardiográfica inicial para eliminar las variaciones de la línea base.

Filtros variantes en el tiempo.

En este caso la frecuencia de corte va variando según el análisis de las componentes frecuenciales de cada latido respecto a un promedio de los mismos. Como ejemplo de este caso se puede considerar el trabajo de [Sör93]. En éste se parte de la expresión de un filtro patrón, el cual puede cambiar su frecuencia de corte de forma simple variando alguno de sus parámetros. Para seleccionar dicha frecuencia de corte se presentan dos métodos. En un primer método se adapta la frecuencia de corte al nivel de línea base presente en la señal, que ha sido estimada a partir del error entre la salida del filtro actual, y la salida producida por el filtro de frecuencia de corte máxima. En el segundo método, la frecuencia de corte se estima a partir de la longitud del intervalo RR de entre 5 y 10 latidos anteriores al latido bajo estudio.

Estructuras de celosías.

Al igual que en la solución propuesta para la eliminación de la frecuencia de red, en este caso se propone en [RV03] la utilización de una estructura de filtros dispuestos en celosía para eliminar y corregir las variaciones de la línea de base para lo que se utilizan las *Ecuaciones 3.17, 3.18 y 3.19*.

3.2. Extracción de características

La extracción de las características de un ECG supone una forma de comprimir la señal eliminando la información redundante y manteniendo todas aquellas propiedades relevantes sin perder información crítica. Para ello normalmente se utilizan métodos que aproximan la señal tratada simplificándola en mayor o menor grado. En la literatura especializada se pueden encontrar un gran número de algoritmos de aproximación de señales biomédicas, como por ejemplo la utilización de polinomios de grado N (con $N > 1$) para el reconocimiento de las ondas en el pulso carotídeo [SKK76] y su compresión [PJ92], la aproximación de la señal electroencefalográfica utilizando funciones de autocorrelación [MH79] o la aproximación de la señal mediante líneas poligonales [KJ96]. En el caso de la extracción de las características en una señal ECG utilizaremos el método de aproximación poligonal con lo que, además de caracterizar la forma de onda se consigue una importante reducción del ruido. Precisamente por este motivo, el autor propone una extracción de características previa a la etapa de segmentación, procesando la señal inmediatamente después de eliminar la interferencia de la red y una vez corregida la línea de base (*Figura 3.8*). De esta forma, además de extraer las características filtrando al mismo tiempo el ruido, se consigue una reducción sustancial de la información tratada con lo que el posterior proceso de segmentación se verá beneficiado al disminuir su carga computacional. Por ejemplo, durante el proceso de segmentación de la señal electrocardiográfica y para la detección del complejo *QRS*, en [VGL97] se realiza inicialmente una aproximación poligonal de la onda con lo cual, además de reducir el ruido, se disminuye la cantidad de datos a tratar.

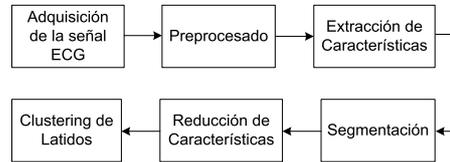


Figura 3.8: La extracción de características es realizada mediante aproximación poligonal, donde la misma etapa de extracción se anticipa siendo utilizada para el filtrado del ruido y la reducción de la información a tratar en procesos posteriores.

3.3. Segmentación

En cualquier tipo de análisis de la señal electrocardiográfica es muy importante la detección de ciertos elementos para obtener la duración y amplitud de las ondas con el fin de aislar los latidos para su posterior clasificación y diagnóstico sobre los resultados obtenidos. Por ejemplo, si la contracción de la aurícula es más lenta de lo normal, el segmento *PQ* se hace más largo y la duración del complejo *QRS* también. Examinando los intervalos *RR* se puede obtener el ritmo cardíaco. Cambios en el tamaño de la onda *T* y/o deformaciones en esta onda o en la pendiente del *ST* pueden ser asociados con arritmias e infartos. En los últimos años se han desarrollado muchos métodos para llevar a cabo un análisis automático de la señal electrocardiográfica [Lag90], donde el principal problema radica en el reconocimiento de sus subpatrones más importantes.

3.3.1. Detección de las ondas *P* y *T*

Corresponde a un caso de detección de puntos de cierta complejidad debido a la poca amplitud de estas ondas, lo cual hace que en muchas ocasiones estén enmascaradas por el ruido.

En [GFP89] se propone un método para llevar a cabo esta tarea utilizando una

transformada longitudinal del ECG.

En [ARD94] se describe un método basado en la transformada *Wavelet* para localizar de forma muy precisa la onda *P*.

En [YHS97] el método descrito utiliza una red neuronal, aunque en este caso se extrae más información: duración de la onda *P*, intervalo *PR*, duración del complejo *QRS*, duración de las ondas *S* y *T*, intervalo *QT*, amplitud de las ondas *P*, *R* y *T*, nivel del segmento *ST*, y área del intervalo *QT*.

En [VFLC95] se utilizan umbrales en lugar de redes neuronales para detectar las ondas *P* y *T*, además del complejo *QRS*.

Finalmente en [AMO⁺03] se presenta un método multiescala para detectar e identificar las ondas *P* y *T* dentro de un amplio rango de morfologías utilizando la ventaja que la transformada *wavelet* presenta en el escalado de la señal en el dominio temporal. El algoritmo presentado tiene una sensibilidad del 98.87% para las ondas *P* y del 99.77% para las ondas *T*. Para ello se utiliza como prototipo de *Wavelet* el spline originalmente propuesto en [LZT95] y en [BHH97]:

$$\Psi(\Omega) = j\Omega \left(\frac{\sin(\Omega/4)}{\Omega/4} \right)^4 \quad (3.31)$$

Para este prototipo, los filtros $H(z)$ y $G(z)$ que se implementan en la *Figura 3.9* son:

$$H(e^{j\omega}) = e^{j\omega/2} \left(\cos \frac{\omega}{2} \right)^3 \quad (3.32)$$

$$G(e^{j\omega}) = 4je^{j\omega/2} \left(\sin \frac{\omega}{2} \right) \quad (3.33)$$

3.3.2. Detección del complejo *QRS*

Dentro de la detección de puntos significativos, el caso más relevante es la detección del complejo *QRS*. Generalmente, para detectar este complejo característico en una señal electrocardiográfica, existen algoritmos que marcan de alguna manera la posición de inicio o fin del complejo *QRS* (*Figura 3.10*).

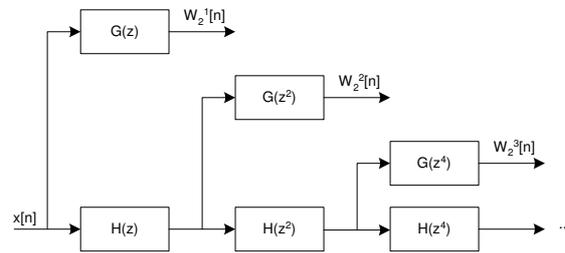


Figura 3.9: Banco de filtros para la detección de las ondas P y T (algorithme à trous).

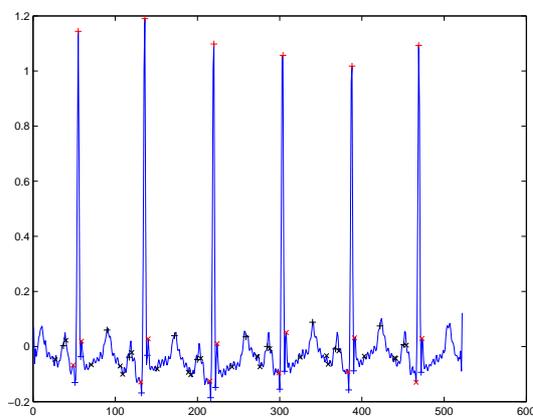


Figura 3.10: Detección del complejo QRS en una señal ECG real. La fuente utilizada procede de la base de datos del MIT-BIH *Arrhythmia Database*.

Para evaluar las prestaciones de cada uno de estos algoritmos, se puede utilizar una medida del número de falsas detecciones realizadas y del número de detecciones falladas. De cualquier forma, estos parámetros sólo tendrán valor comparativo cuando se hayan aplicado a las mismas señales, ya que según sean las características de éstas, las prestaciones de los algoritmos varían. En [FJJ⁺90] se realiza un análisis comparativo de varios algoritmos según su sensibilidad al ruido, además de clasificarlos genéricamente en una serie de algoritmos (los basados en el tratamiento digital de señales, los basados en la amplitud y en la primera derivada, los basados únicamente en la primera derivada, los basados en la primera y segunda derivada, los basados en filtros

digitales, los basados en la comparación de patrones y los basados en transformaciones no lineales), que se describen a continuación:

- *Algoritmos basados en el tratamiento digital de señales*, que utilizan procedimientos clásicos de teoría de la señal [CAP99] [DPH93]. Debido a su generalidad, en la aplicación a señales electrocardiográficas en concreto, son superados por muchos otros, por lo que no se recurre a ellos en la práctica. Entre las técnicas y procedimientos más comunmente empleados cabe citar el de *Autocorrelación*, la técnica *AMDF* (*Average Mean Distance Function*) y la *Transformada Rápida de Fourier* (*Fast Fourier Transform*, FFT).

La *autocorrelación* de una señal discreta $x[n]$ se expresa matemáticamente como:

$$r_x[\eta] = \varepsilon\{x[n]x[n - \eta]\} \quad (3.34)$$

donde, en el caso de que $x[n]$ sean las muestras de un proceso ergódico, tiene la expresión temporal:

$$r_x[n] = \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \sum_{n=-N}^N x[n]x[n - \eta] \quad (3.35)$$

Esta función presenta máximos en aquellos puntos η donde dos traslaciones de $x[n]$ tienen mayor parecido. De esta forma, aplicando esta expresión a la señal electrocardiográfica, es de suponer que los máximos indicarán el periodo de la señal. Sin embargo, debido al ruido y la variabilidad inherente a este tipo de señales, existen zonas en las que aparecen varios máximos, lo cual hace difícil la localización exacta del máximo que mejor define el periodo, por lo que este método no resulta apropiado [CAP99]. Hay algunas modificaciones aplicadas a esta función para mejorar las prestaciones, como elevar la señal a cierta potencia, realizar un filtrado inicial, o aplicar la técnica de *Center Clipping* [DPH93].

Para el cálculo de la frecuencia fundamental de la señal se puede utilizar la técnica *AMDF*, que resulta similar a la autocorrelación con la diferencia que en este caso se emplea la resta de dos traslaciones de la misma señal en lugar de su producto. Su expresión matemática es [DPH93]:

$$\Delta M_x[n] = \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \sum_{n=-N}^N |x[n] - x[n - \eta]| \quad (3.36)$$

En este caso, para localizar una repetición, se ha de buscar un mínimo en la señal resultante. Sin embargo, este método presenta inconvenientes similares al caso de la autocorrelación ya que en algunas zonas aparecen varios mínimos y es difícil distinguir cuál de ellos es el más apropiado. En cuanto a la *Transformada Rápida de Fourier*, se basa en el cálculo de la transformada discreta de Fourier de la señal [OS89]. Si la señal fuera perfectamente periódica aparecerían pulsos en el espectro localizados en múltiplos enteros de la frecuencia fundamental de la señal. En este caso, no es perfectamente periódica aunque se puede apreciar un efecto similar al descrito, tanto aplicando la FFT a toda la señal, como realizando previamente un inventariado rectangular de la señal. Este efecto se puede realzar aplicando el espectro de productos armónicos (HPS) [DPH93]. Se trata de llevar a cabo el producto de varios espectros cuya componente frecuencial ha sido escalada por una componente entera, de forma que se consigue realzar la frecuencia fundamental, ya que esta técnica es bastante resistente al ruido. El problema planteado por los algoritmos basados en la FFT es que sólo indican la duración del periodo, pero no su comienzo ni su final, lo cual no resultan demasiado útiles para esta aplicación.

- *Algoritmos basados en la amplitud y en la primera derivada*, referentes a aquellos métodos que detectan la presencia del complejo QRS basándose en la aplicación de umbrales sobre la amplitud ($x[n]$) y primera derivada ($y[n] = x[n+1] - x[n-1]$) de la señal [FJJ⁺90]. De esta manera, y ajustando los parámetros del algoritmo al tipo de señal a tratar, se consigue identificar el complejo QRS al comparar los valores de amplitud y derivada con el umbral establecido. De todas formas, hay que tener en cuenta que, en general, los resultados ofrecidos por estos algoritmos no son satisfactorios si la señal no se encuentra suficientemente limpia de artefactos como el ruido y las variaciones de la línea base.
- *Algoritmos basados únicamente en la primera derivada*, donde, al contrario que en el caso anterior, se omite la amplitud como elemento de decisión. Dentro de este grupo estaría el algoritmo planteado por *Menard* y descrito en [FJJ⁺90] donde, para la detección del complejo QRS, se realiza la comparación de la primera derivada de la señal ($y[n] = -2x[n-2] - x[n-1] + x[n+1] + 2x[n+2]$) con cierto umbral.

- *Algoritmos basados en la primera y segunda derivada*, métodos propuestos por Balda y descritos también en [FJJ⁺90], donde se calculan la primera y segunda derivadas de la forma:

$$y_1[n] = \text{abs}(x[n+1] - x[n-1]) \quad (3.37)$$

$$y_2[n] = \text{abs}(x[n+2] - 2x[n] + x[n-2]) \quad (3.38)$$

Posteriormente se ponderan las dos señales y se suman:

$$y_3[n] = 1,3y_1[n] + 1,1y_2[n] \quad (3.39)$$

Finalmente se compara con cierto umbral para identificar los puntos del complejo QRS.

Otro de los algoritmos utilizados en este punto sería el de *Ahlstrom* y descrito en [FJJ⁺90].

- *Algoritmos basados en filtros digitales*, un algoritmo de este tipo fue propuesto por *Zeelenberg* [FJJ⁺90]. Consiste en el filtrado de la señal a través de un filtro *Notch* y por un diferenciador de la forma $y[n] = x[n] - x[n-4]$ y filtrando paso bajo mediante la ecuación:

$$y_1[n] = y[n] + 4y[n-1] + 6y[n-2] + 4y[n-3] + y[n-4] \quad (3.40)$$

comparando posteriormente con dos umbrales para identificar el complejo QRS. En [PTCJ90] se describe un filtro para la detección tanto del complejo QRS como para la detección de las ondas *P* y *T*. Este filtro está relacionado con los algoritmos basados en la primera derivada, aunque incluye una parte de filtrado paso bajo. En realidad, este filtro viene a fusionar un filtro paso bajo y un derivador en un solo filtro.

- *Algoritmos basados en la comparación de patrones* de cada una de las ondas o de algunos complejos como el QRS contra ciertas partes del electrocardiograma. El trabajo descrito en [VGL97] utiliza la comparación de las ondas mediante alineamiento temporal. Para ello, y en primer lugar, realiza una aproximación de la onda según el método propuesto por *Koski* [KJM95], con lo cual, además de reducir el ruido, se disminuye la cantidad de datos. Seguidamente, se utiliza otro método para la detección del complejo QRS y se establece cada periodo como el intervalo entre dos ondas *R* consecutivas. A continuación se toma una onda

como referencia y se compara al resto mediante alineamiento temporal usando programación dinámica (Dynamic Time Warping, DTW) para obtener la disimilitud entre ellas. Sin embargo, en este trabajo sólo consigue el camino de alineamiento (basándose en la pendiente de las rectas aproximadas) para decidir la correspondencia entre las ondas tomadas como referencia y las obtenidas del electrocardiograma. Finalmente, el autor establece la necesidad de realizar mejoras al procedimiento debido a que los resultados no son suficientemente satisfactorios. Otros métodos basados en patrones se fundamentan en el reconocimiento sintáctico [KJM95] [TS90] [TS89]. Para ello, descomponen el electrocardiograma en sus ondas componentes, lo cual implica la extracción de primitivas, representación lingüística y formulación en forma de gramática. El problema asociado a estos métodos es el del coste computacional relativamente elevado que conllevan, además de los errores que se producen debido a la inexactitud de las etapas iniciales de los algoritmos.

- *Algoritmos basados en transformaciones no lineales*, en este caso se proponen las referencias [SS94] a tres de los algoritmos utilizados: algoritmo MOBD (*Multiplication Of Backward Difference*), algoritmo de Okada y algoritmo de *Hamilton-Tompkins*. La comparativa de los resultados obtenidos para el procesado de un total de 358.551 latidos, para cada uno de los tres casos, es la que se muestra en el Cuadro 3.2

Cuadro 3.2: Comparativa de tres algoritmos de detección del complejo QRS basados en transformaciones no lineales.

Algoritmo	Bits de Datos	FP	FN	% Error
MOBD	7	11633	13253	6.94
Okada	12	13066	26510	11.04
Hamilton-Tompkins	12	6746	7068	3.85

- *Algoritmos basados en la transformada wavelet*, donde, por ejemplo, se utiliza la potencia de wavelets ortogonales para el filtrado y detección del complejo QRS en señales ECG [ARCG05]. En [Joh04] se propone la identificación de los complejos QRS utilizando la descomposición en wave-

lets de un nivel, a diferencia de los métodos desarrollados hasta el momento [DKPB01], [KMBB99], [LZT95] o [Mal91].

3.4. Reducción de características

La etapa de reducción o selección de las características de un objeto resulta ser el paso más importante dentro de un proceso global de agrupamiento. En esta etapa se pretenden extraer de los datos el menor número de características que nos proporcione la mayor cantidad de información y que nos permita discriminarlos correctamente durante la posterior fase de clasificación. Si se seleccionan características con poca capacidad de discriminación, los resultados de la clasificación resultarán confusos. Por el contrario, si se consigue caracterizar los datos conservando sus rasgos diferenciales, el posterior proceso de clustering se simplifica enormemente y los resultados mejoran respecto al caso anteriormente expuesto [TK99]. En la literatura relativa a este tema aparecen una gran cantidad de transformaciones a aplicar sobre los datos [CNPA02]: la transformada de Karhunen-Loeve [MM89], las transformadas Wavelet [LZT95], la transformada de Fourier; el análisis estadístico multivariante: análisis de componentes independientes (Independent Component Analysis, ICA) [HO00], análisis de componentes principales (PCA), análisis multivariante de la varianza; o incluso otros métodos utilizados normalmente para la caracterización de datos: momentos, modelado paramétrico, modelos ocultos de Markov (HMM), etc. En el caso que nos ocupa y en esta etapa, utilizaremos dos métodos distintos para reducir las características de los latidos de una señal ECG: (i) el PCA y (ii) los HMM.

3.4.1. Análisis de componentes principales

El PCA utilizado en la etapa de selección o reducción de las características en señales ECG se engloba dentro de las técnicas de análisis estadístico multivariante. Este tipo de análisis estadístico se caracteriza por englobar una amplia gama de métodos basados en el análisis de los datos considerando sus variables en conjunto en lugar de centrar los esfuerzos de análisis sobre sus variables individuales. Afortunadamente y en conjuntos de datos con un

número de variables elevado, es norma común el que las variables varíen de forma grupal y no individualmente. Esto es así ya que normalmente algunas de las variables actúan o están relacionadas con un mismo principio fundamental que gobierna el sistema que se pretende analizar. Es decir, las técnicas de medida actuales nos proporcionan información redundante al permitirnos evaluar un elevado número de parámetros. El análisis multivariante trata de simplificar los sistemas eliminando la información redundante de un grupo de variables al combinarlas en una única variable nueva.

Uno de estos métodos multivariantes es el que aborda el análisis sobre las componentes principales de los datos [Mat01] y que consiste en la búsqueda de una base que maximice la varianza entre los datos proyectados sobre ésta. El conjunto de las nuevas variables generadas se denomina *componentes principales* del sistema. Cada una de las componentes principales es una combinación lineal de las variables originales que, además, es ortogonal al resto de componentes, de manera que se elimina cualquier tipo de redundancia. Las componentes principales conforman una base ortogonal del conjunto de datos analizados (vectores propios) y vienen ponderadas por un factor de peso dependiendo de su importancia en el espacio final (valores propios).

Realmente y para un número determinado de variables existen infinitas bases ortogonales, por lo que la pregunta sería ¿qué ventaja supone la base ortogonal constituida por las componentes principales respecto del resto de bases?

La primera de las componentes principales define un eje en el espacio n -dimensional (donde n es el número de variables que definen el modelo a analizar). La nueva variable creada al proyectar cada una de las observaciones sobre el eje definido por la primera componente principal, se caracteriza por tener una varianza máxima (de entre todas las varianzas calculadas para las posibles variables resultado de la proyección de las observaciones sobre cualquier otro eje seleccionado).

La segunda de las componentes principales representa otro eje en el espacio (ahora bidimensional) perpendicular al primero. La proyección de las observaciones sobre este segundo eje proporciona una nueva variable cuya varianza vuelve a ser la máxima de las posibles para cualquier otro de los ejes ortogonales que hubieran podido ser elegidos.

El conjunto de componentes principales coincide con el número de variables en el espacio original (n) presentando como ventaja que, al maximizar la va-

rianza, mediante un reducido número de componentes principales (las 3 ó 4 primeras) se aglutinan porcentajes de la varianza total de los datos originales de hasta un 80%. De esta forma se consigue eliminar una gran cantidad de información redundante lo que facilita a los analistas la interpretación de los datos (incluso de manera visual, si sólo se representan tridimensionalmente las tres primeras componentes).

Matemáticamente y suponiendo que X es la matriz de datos (centrados) a analizar mediante PCA, vamos a buscar un vector unitario u que maximice la varianza λ de la proyección de las columnas de X (variables) sobre el eje definido por dicho vector. Esto nos lleva a diagonalizar la matriz $X' \cdot X$ [AM99]. Esta matriz es igual a la matriz de correlaciones entre variables en el caso de un PCA normalizado y es igual a la matriz de covarianzas entre variables para el caso de un PCA no normalizado. El vector unitario buscado u , es el vector propio asociado al mayor valor propio obtenido en la diagonalización de $X' \cdot X$.

Análogamente, la dirección ortogonal al vector u , y que a su vez maximiza la inercia proyectada sobre ella, corresponde al vector propio asociado al segundo valor propio obtenido al diagonalizar la matriz $X' \cdot X$, siendo esta inercia proyectada igual al segundo valor propio y así sucesivamente.

Los valores propios dan directamente las inercias proyectadas sobre cada una de las direcciones que buscamos y cuya suma es la suma de la inercia sobre las direcciones ortogonales, que es igual a la inercia global de la nube de puntos originales.

Los vectores propios dan las direcciones de máxima inercia y los denominamos ejes factoriales.

El PCA consiste en pasar de las n variables originales x_j , cada una con una importancia medida por su varianza, a n nuevas variables Ψ_α , combinaciones lineales de las originales (Ecuación 3.41), cada una con una importancia medida por su varianza, igual a su valor propio (Ecuación 3.42).

$$\Psi_\alpha = u_1x_1 + \dots + u_nx_n \quad (3.41)$$

$$var(\Psi_\alpha) = \lambda_\alpha \quad (3.42)$$

A pesar de la sencillez del método propuesto, se nos presentan dos pro-

blemas importantes: la técnica carece de un modelo probabilístico definido y, además, es globalmente lineal. Para solucionar estas limitaciones se propone un modelo de mezclas denominado análisis probabilístico de componentes principales (PPCA) donde se define un modelo probabilístico del PCA que simula los datos de manera no lineal mediante una mezcla de submodelos lineales locales [TB99].

3.4.2. Modelos ocultos de Markov

La utilización de HMM para el agrupamiento de los latidos en una señal ECG plantea el dilema de su identificación en alguna de las etapas que componen el proceso global de clustering. Esto es así ya que, por un lado, podemos considerar el modelado de los objetos como una reducción de sus características a un índice probabilístico generado por el modelo en cuestión: un latido se identifica con un modelo si la probabilidad de la secuencia de segmentos sobre el modelo supera cierto umbral y el vector de características del objeto queda reducido a un vector de probabilidades de longitud igual al número de modelos identificados. El problema es que, por otro lado, y para completar el clustering con HMM se necesita generar un modelo por cada una de las distintas morfologías presentadas en el ECG, con lo que el cálculo de la matriz de similitud depende de la acertada selección de los centroides y de un correcto entrenamiento de los modelos iniciales. Es decir, el conjunto de HMMs proporcionan la matriz de similitud sobre la que aplicar el algoritmo de clustering, pero no conocemos ese conjunto de modelos hasta identificar adecuadamente los centroides (para lo cual es necesario aplicar a su vez algún algoritmo de clustering). Aunque la frontera no está clara, para simplificar la redacción de este documento se ha preferido considerar los modelos como una forma de reducir las características. La solución final al clustering con HMM es la que se propone en la *Figura 8.1* del *Capítulo 8*.

A continuación se presenta la teoría de los HMM y de todos aquellos desarrollos y aplicaciones relacionados con la misma. Inicialmente, y con relación a las técnicas de reconocimiento de formas (modelado y comparación de patrones), se presentarán someramente los diferentes métodos y algoritmos existentes en la literatura para pasar a describir la teoría de los HMM incluyendo, para su mejor comprensión, una reseña sobre los modelos de mezclas de distribución

de probabilidad gaussiana (Gaussian Mixture Models, GMM) y otra relativa a las cadenas de Markov. Después revisaremos los diferentes algoritmos de entrenamiento no supervisado utilizados en los HMM. También se incluye una sección donde se presentan todas aquellas mejoras desarrolladas para incrementar las prestaciones de este tipo de modelos. Finalmente se incluye un punto que documenta algunas de las aplicaciones implementadas y que se basan en los HMM.

Modelos de mezclas con distribución de probabilidad gaussiana

Los modelos de mezclas y en particular los que se basan en mezclas de fuentes gaussianas, han constituido desde siempre una herramienta comúnmente utilizada en operaciones de estimación de densidades de probabilidad, de clustering y de entrenamiento no supervisado así como en aplicaciones del campo de la estadística, reconocimiento de patrones [Moe00] y data mining [DHS01] [Bis95] [MP00], además de constituir una herramienta muy potente para el tratamiento de conjuntos de datos y, en particular, de variables ocultas. Si, dado un modelo basado en fuentes gaussianas, la variable oculta es la que indica la fuente gaussiana que la genera; el modelo global quedará perfectamente definido por la función de distribución de probabilidad conjunta formada por la función de densidad de probabilidad de la variable observada más la de la variable oculta.

Sea $\mathbf{X} = [X_1, \dots, X_d]$ una variable aleatoria d -dimensional, siendo $\mathbf{x} = [x_1, \dots, x_d]$ una realización concreta de \mathbf{X} . Resulta que \mathbf{X} sigue una distribución de mezclas finitas con K -componentes si su función de densidad de probabilidad se puede expresar como:

$$p(\mathbf{x} | \Theta) = \sum_{k=1}^K p(k)p(\mathbf{x} | \Theta_k) \quad (3.43)$$

donde $p(1), \dots, p(K)$ son las funciones de densidad de probabilidad que componen la mezcla, Θ_k es el conjunto de parámetros que definen la k -ésima componente de la mezcla, y $\Theta = \{\Theta_1, \dots, \Theta_k, p(1), \dots, p(K)\}$ es el conjunto de parámetros que se necesitan para definir completamente la mezcla. En este

punto, supondremos que todas las componentes presentan una distribución gaussiana de la forma mostrada en la Ecuación 3.44, con cada componente caracterizada por $\Theta_k = \{\mu_k, \sigma_k\}$. Además, también se asumirá independencia estadística entre los vectores de características (definimos el n -ésimo vector de características que representa al n -ésimo objeto como $\mathbf{x}^n = [x_1^n, \dots, x_d^n]$) para que, de esta forma la matriz de covarianzas derive en una matriz (vector) de varianzas σ .

$$p(\mathbf{x}^n | \Theta_k) = \prod_{i=1}^d \frac{1}{\sqrt{2\pi}\sigma_{k,i}} \exp -\frac{(x_i^n - \mu_{k,i})^2}{2\sigma_{k,i}^2} \quad (3.44)$$

Por lo tanto, y dado un conjunto de N muestras independientes $\mathcal{X} = \{\mathbf{x}^1, \dots, \mathbf{x}^N\}$, la función de probabilidad logarítmica que corresponde a la k -ésima componente de la mezcla resulta:

$$P(\mathcal{X} | \Theta) = \mathcal{L}(\Theta | \mathcal{X}) = \log \prod_{i=1}^N p(\mathbf{x}^i | \Theta) = \sum_{k=1}^K p(k) p(\mathbf{x}^i | \Theta_k) \quad (3.45)$$

El método estándar utilizado para el modelado de los datos mediante este tipo de mezclas es el algoritmo de *expectation-maximization* (EM) que trata de estimar los parámetros del modelo de forma que se va convergiendo hacia el punto de máxima probabilidad (maximum likelihood o ML) [MK97a]. Las sucesivas aproximaciones de los parámetros que refinan el modelo de mezclas de fuentes gaussianas se pueden calcular mediante la iteración de las ecuaciones para todo $k \in \{1, \dots, K\}$

$$p(k | \mathbf{x}^i) = \frac{p(k)p(\mathbf{x}^i | \Theta_k)}{P(\mathcal{X} | \Theta)} \quad (3.46)$$

$$p(k) = \frac{1}{N} \sum_{i=1}^N p(k | \mathbf{x}^i) \quad (3.47)$$

$$\mu_k = \sum_{i=1}^N \frac{p(k | \mathbf{x}^i)x^i}{Np(k)} \quad (3.48)$$

$$\sigma_k = \sum_{i=1}^N \frac{p(k | \mathbf{x}^i)(\mathbf{x}^i - \mu_k)^2}{Np(k)} \quad (3.49)$$

Cadenas de Markov

Las *Cadenas de Markov* se utilizan normalmente para modelar aquellos procesos aleatorios que requieren de cierta memoria. Sea S_1, S_2, \dots, S_N una secuencia de variables aleatorias cuyos valores se representen mediante un alfabeto de símbolos finito $\mathcal{X} = \{1, 2, \dots, c\}$, y aplicando la fórmula de Bayes

$$P(S_1, S_2, \dots, S_N) = \prod_{i=1}^N P(S_i | S_1, S_2, \dots, S_{i-1}) \quad (3.50)$$

Se dice que las variables aleatorias forman una cadena de Markov [Jel99] si:

$$P(S_i | S_1, S_2, \dots, S_{i-1}) = P(S_i | S_{i-1}) \quad \forall i \quad (3.51)$$

Y como consecuencia, en una cadena de Markov:

$$P(S_1, S_2, \dots, S_N) = \prod_{i=1}^N P(S_i | S_{i-1}) \quad (3.52)$$

En este caso, los procesos aleatorios tienen una capacidad de memoria tan limitada que el valor de la variable en el instante de tiempo t depende únicamente del valor inmediatamente anterior y de ningún otro. Las cadenas de

Markov serán invariantes en el tiempo si a pesar del valor del índice temporal i se cumple:

$$P(S_i = s' | S_{i-1} = s) = p(s' | s) \quad \forall s, s' \in \mathcal{X} \quad (3.53)$$

donde $p(s' | s)$ resulta ser la función de transferencia que puede tomar forma de matriz $c \times c$. Además, $p(s' | s)$ debe satisfacer las condiciones típicas para todo $s \in \mathcal{X}$

$$\sum_{s' \in \mathcal{X}} p(s' | s) = 1, \quad p(s' | s) \leq 1, \quad s' \in \mathcal{X}. \quad (3.54)$$

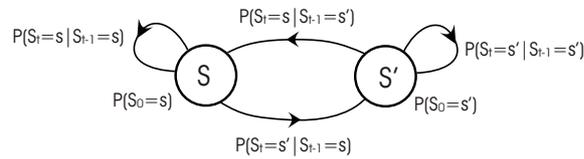


Figura 3.11: Cadena de Markov de dos estados.

Podríamos suponer que S_i son los *estados* con lo que la cadena de Markov resultaría un proceso de estados finitos cuyas transiciones entre los estados vendrían definidas por la función de transferencia $p(s' | s)$. Un ejemplo de cadena de Markov de 2 estados ($c = 2$) es la que se muestra en la *Figura 3.11*.

Modelos Ocultos de Markov

Como ya hemos comentado anteriormente, en el presente trabajo se utilizará un reconocedor de formas estocástico (HMM) como base para el reconocimiento y clasificación de los latidos de un ECG Holter. Se podría definir un HMM como un autómata de estados finitos y estocástico caracterizado por los siguientes parámetros [Rab89]:

1. Número de estados del modelo, N . Aunque los estados permanecen ocultos, para muchas aplicaciones prácticas se pueden relacionar éstos

con alguna magnitud o característica física. Nos referimos a la secuencia de estados que conforman el modelo como $S = (S_1, S_2, \dots, S_N)$, y al estado en el instante de tiempo t como q_t .

2. Número de símbolos distintos por estado M , o número de fuentes gaussianas que participan en función de densidad de probabilidad conjunta.
3. Matriz de probabilidad de transición entre los estados $A = \{a_{ij}\}$, de tamaño $N \times N$, y que define la probabilidad que existe de encontrarnos en el estado i en el instante de tiempo t , para pasar a encontrarnos en el estado j en $t + 1$.

$$a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i), \quad 1 \leq i, j \leq N. \quad (3.55)$$

4. Probabilidades iniciales de los estados $\pi = \{\pi_i\}$, donde se determinan las probabilidades de cualquiera de los estados en el instante inicial, donde:

$$\pi_i = P(q_1 = S_i), \quad 1 \leq i \leq N. \quad (3.56)$$

5. Probabilidades de emisión, que podemos dividir en dos categorías dependiendo de si la secuencia observada es de tipo discreto o continuo:

- Probabilidad de emisión discreta $B = \{b_j(k)\}$ por lo que, si M son los diferentes símbolos observados en cada estado (es decir, el tamaño del alfabeto utilizado), tendremos un alfabeto $V = \{v_1, v_2, \dots, v_M\}$. Así pues, b_j sería la probabilidad de observación del símbolo si nos encontramos en el estado j

$$b_j(k) = P(v_k \text{ at } t \mid q_t = S_j) \quad 1 \leq j \leq N, 1 \leq k \leq M \quad (3.57)$$

La secuencia de símbolos observados constituye la salida que se obtiene del sistema a modelar. Este tipo de sistemas se llama HMM discretos (Figura 3.12).

- Probabilidad de emisión continua $B = \{b_j(O_t)\}$, donde $\mathbf{O} = O_1, O_2, \dots, O_T$. La función de densidad de probabilidad de emisión para cada estado se define como una mezcla finita de fuentes gaussianas (3.58):

$$b_j(O_t) = \sum_{m=1}^M d_{jm} \mathcal{N}(O_t, \mu_{jm}, C_{jm}), \quad 1 \leq j \leq N \quad (3.58)$$

donde O_t es el vector de características de la secuencia de observaciones a modelar, d_{jm} son los coeficientes de la m -ésima mezcla del estado j y \mathcal{N} es una probabilidad gaussiana con vector de media μ_{jm} y matriz de covarianzas C_{jm} correspondiente a la m -ésima componente de la mezcla en el estado j . estos modelos se conocen como HMM continuos (CHMM).

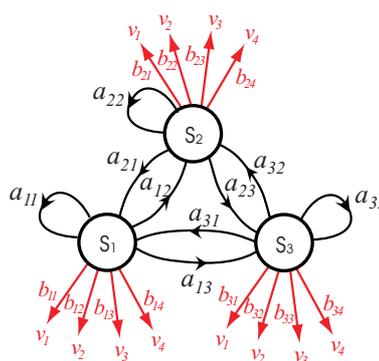


Figura 3.12: HMM discreto de tres estados con cuatro símbolos por estado.

Para la completa especificación de un HMM necesitaremos determinar los dos parámetros que determinan la *geometría* del modelo N y M , y las tres probabilidades A, b, π . Todo ello lo hacemos mediante la notación:

$$\lambda = \{A, B, \pi\} \tag{3.59}$$

Otro de los aspectos importantes de esta teoría es la de las restricciones que se derivan de un modelo estocástico, a saber:

$$\begin{aligned}
\sum_{i=1}^N \pi_i &= 1 \\
\sum_{j=1}^N a_{ij} &= 1, \quad 1 \leq i \leq N \\
\sum_{k=1}^M b_j(k) &= 1, \quad 1 \leq j \leq M
\end{aligned} \tag{3.60}$$

Así pues, y una vez definido el HMM, se nos plantean tres problemas inmediatos cuya solución se puede encontrar en [Rab89]:

1. Dada la secuencia observada $\mathbf{O} = O_1, O_2, \dots, O_T$ y el modelo λ , ¿cómo calcular la probabilidad resultante de dicha secuencia $P(\mathbf{O} \mid \lambda)$, dado el modelo anterior [Rab89]? Este problema se conoce con el nombre de *inferencia*.
2. Dada la secuencia de observaciones $\mathbf{O} = O_1, O_2, \dots, O_T$ y el modelo λ , ¿cómo encontrar la secuencia de estados (ocultos) $Q = q_1, q_2, \dots, q_T$ que mejor explique la secuencia observada? Este problema se resuelve mediante el algoritmo de Viterbi [Vit67].
3. ¿Cómo ajustar los parámetros del modelo $\lambda = \{A, b, \pi\}$ para maximizar $P(\mathbf{O} \mid \lambda)$? Este problema se conoce con el nombre de *aprendizaje*.

Algoritmos de aprendizaje en los HMM

En esta sección se van a revisar los métodos de entrenamiento no supervisado existentes en la literatura para su aplicación sobre los HMM. Los podemos clasificar en dos grupos: (i) *algoritmos de optimización o búsqueda ascendente* (del inglés hill-climbing, como por ejemplo, EM, k-medias segmentado y el de búsqueda del gradiente) y (ii) *algoritmos de búsqueda global* (algoritmos genéticos y de *simulated annealing*).

Los algoritmos de búsqueda ascendente dependen enormemente de la manera en la que se inicialice el modelo, de tal forma que, en la práctica y si los parámetros iniciales no han sido los óptimos, la búsqueda puede conducir a

un modelo sub-óptimo. Para evitar este problema se proponen una serie de técnicas [JR90] aunque éstas impliquen una mayor carga computacional. Por otra parte, los algoritmos de búsqueda global no dependen en exceso de la inicialización del modelo, precisamente por su capacidad global para encontrar el óptimo.

A continuación se reseñan todos aquellos algoritmos de entrenamiento que se consideran más relevantes:

- **Algoritmo de Baum-Welch.** El algoritmo EM (*Expectation-Maximization*) [DLR77] es un método general que se utiliza para estimar los parámetros del modelo de tal forma que se maximice la probabilidad (*maximum-likelihood*, ML) de una distribución generada a partir de un conjunto incompleto de datos observados (existen datos no conocidos por algún motivo). Podemos encontrar principalmente dos aplicaciones del algoritmo EM: la primera cuando, efectivamente, el conjunto de datos observados resulta incompleto; la segunda, cuando la optimización de la función de probabilidad es demasiado compleja y se necesitan asumir ciertas simplificaciones (similares a la pérdida de información) para resolver el problema de optimización. Esta última aplicación se utiliza normalmente en tareas de reconocimiento de patrones [Bil98]. El problema de maximización de la función de probabilidad planteado es el siguiente: sea $p(x | \Theta)$ la función de densidad de probabilidad parametrizada por el conjunto de valores definidos en Θ (por ejemplo, p podría ser un conjunto de fuentes gaussianas y Θ los valores de las medias y las varianzas que las definen); y sea el conjunto de datos de tamaño N generados por la distribución anterior $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$. Asumiendo la independencia de las observaciones, la función de densidad de probabilidad de los datos observados resulta:

$$p(\mathcal{X} | \Theta) = \prod_{i=1}^N p(x_i | \Theta) = \mathcal{L}(\Theta | \mathcal{X}) \quad (3.61)$$

función de probabilidad que depende de los parámetros en Θ . El objetivo del algoritmo EM consiste en encontrar los valores de Θ que maximicen nuestra función \mathcal{L} (Ecuación 3.62).

$$\Theta^* = \underset{\Theta}{\operatorname{argmax}} \mathcal{L}(\Theta | \mathcal{X}) \quad (3.62)$$

Normalmente, y para simplificar los cálculos y evitar problemas de overflow, lo que se hace es maximizar su equivalente logarítmico $\log \mathcal{L}(\Theta | \mathcal{X})$. El desarrollo del algoritmo EM se divide en dos pasos claramente diferenciados: primeramente se realiza una estimación (*expectation*, E) de los parámetros de la función de probabilidad suponiendo que el conjunto de datos estuviera completo y, en segundo lugar, se maximiza la función con los valores de los parámetros supuestos en el paso Anterior (*maximization*, M) [MK97a]. Esta secuencia se repite (el valor de la probabilidad logarítmica aumenta en cada iteración) hasta que se encuentra el máximo local de la función de probabilidad. La aplicación del algoritmo EM sobre los modelos de Markov se puede encontrar en [Bil98] donde, en este contexto, se da a conocer también como el algoritmo de *Baum-Welch* [BP66], [DLR77].

- **Búsqueda del gradiente.** En este caso y, a diferencia del EM, el algoritmo de búsqueda del gradiente trabaja on-line, directamente sobre las muestras obtenidas convergiendo mucho más rápidamente hacia el máximo. Esto es debido a que, mientras que la convergencia del EM no tomaba el camino de mayor gradiente sobre la superficie de la función de probabilidad, necesitando de un número elevado de iteraciones hasta llegar al máximo, para el algoritmo de búsqueda del gradiente se define un ratio de aprendizaje que le permite converger hacia el máximo de manera mucho más rápida. Además, y computacionalmente hablando, las ecuaciones que definen el gradiente negativo sobre la función de probabilidad logarítmica $-\log P(\mathbf{O} | \lambda)$ se pueden derivar directamente en vez de tener que aplicar sobre ellas el operador de Lagrange (tal y como se hace con el algoritmo EM) [BC94], obteniendo de esta manera una sencilla regla para el gradiente de la función que permite un aprendizaje on-line sobre los datos de entrada [SR99].
- **K-medias segmentado (Viterbi).** Mediante este método se tratan de ajustar los parámetros del modelo $\lambda = \{A, B, \pi\}$ para maximizar $P(\mathbf{O}, I | \lambda)$ donde I es la secuencia de estados óptima calculada mediante el algoritmo de Viterbi [Vit67].

Asumiendo que tenemos un conjunto de entrenamiento formado por ω secuencias de T observaciones, podemos concretar el algoritmo de entrenamiento k-medias segmentado en los siguientes pasos:

1. Se seleccionan aleatoriamente y se extraen de las observaciones N símbolos, que definirán N clusters o grupos a los que asociaremos mediante mínima distancia euclídea cada uno de los $\omega \times T$ vectores de observaciones del conjunto. A los clusters obtenidos se les denomina estados del modelo.
2. A partir de los estados obtenidos y de las secuencias observadas, se actualizan los coeficientes a_{ij} (matriz de transición entre estados)
3. A continuación se calculan los parámetros que definirán las probabilidades de emisión de cada estado:

\hat{c}_{jm} = como el porcentaje de vectores del estado j que han sido clasificados en el cluster m

$\hat{\mu}_{jm}$ = como la media de los vectores del estado j clasificados en el cluster m

\hat{U}_{jm} = como la matriz de covarianzas de los vectores del estado j clasificados en el cluster m

4. Ahora se utiliza Viterbi para calcular la secuencia de estado ocultos que mejor explica cada una de las secuencias de entrenamiento mediante: $\hat{\lambda}_i = \{\hat{A}_i, \hat{B}_i, \hat{\pi}_i\}$
5. Finalmente, si alguno de los vectores es reasignado a un nuevo estado en el *Paso 4*, se utiliza la nueva distribución de los clusters para repetir los *Pasos 2 a 5*. Si esto no sucede, el algoritmo termina.

- **Adaptación del teorema de Bayes.** En la literatura clásica, el conjunto de parámetros λ que optimizan el modelo a la vista de los datos observados son desconocidos pero ciertamente existen y tienen un valor concreto. En la aproximación *bayesiana* al problema de optimización, se considera que dicho conjunto de parámetros no tiene un valor fijo sino que éste varía de alguna manera. Esta variación se puede describir mediante una función de probabilidad subjetiva que llamaremos distribución previa. La ventaja de esta función es que puede ser formulada antes de conocer los datos con lo que, para maximizar la función de probabilidad del

conjunto de las observaciones se procede a actualizar λ utilizando el teorema de Bayes sobre su distribución previa. A continuación se reseñan brevemente dos de los métodos basados en la adaptación del teorema de Bayes: la estimación del máximo a posteriori (maximum a posteriori, MAP), y el aprendizaje diferencial (variational learning estimation, VAR).

- *Aprendizaje por estimación del máximo a posteriori (MAP)*: este método incorpora la distribución previa en el proceso de aprendizaje. De esta manera se soluciona el problema por el que se estima erróneamente la función de máxima probabilidad (ML) cuando la cantidad de datos es escasa [GL94]. Precisamente la diferencia entre MAP y ML estriba en la existencia de esa función de distribución previa, que puede ser deducida de la secuencia de observaciones \mathbf{O} como una función de densidad de probabilidad $P(\mathbf{O} \mid \lambda)$, y si $P_0(\lambda)$ es la función de densidad de probabilidad previa correspondiente al conjunto de parámetros λ , entonces la estimación realizada con el MAP resulta:

$$\lambda_{MAP} = \underset{\lambda}{\operatorname{argmax}} P(\lambda \mid \mathcal{O}) \approx \underset{\lambda}{\operatorname{argmax}} P(\mathcal{O} \mid \lambda) P_0(\lambda) \quad (3.63)$$

donde $\mathcal{O} = \{\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_L\}$ es el conjunto de datos observados. Si se asume que λ es desconocido y fijo, no existe ningún conocimiento previo sobre dicho λ o, lo que es lo mismo, la función de distribución previa no nos proporciona ningún tipo de información sobre de los datos $P_0(\lambda) = \text{constante}$. Bajo estas premisas, la Ecuación 3.63 deriva en la conocida formulación de la función ML. Por otra parte, y una vez establecida la formulación del algoritmo MAP, hay que hacer hincapié en tres cuestiones clave: (i) la elección de la familia de funciones de distribución previa, (ii) la definición de los parámetros para la función de densidad de probabilidad previa y (iii) el cálculo del máximo a posteriori. La acertada elección de los parámetros mencionados puede simplificar enormemente el proceso de entrenamiento con MAP [GL94]. Como estimador que mida la convergencia del proceso de training se utiliza el valor de la probabilidad logarítmica que, en este caso, consta de dos partes: la probabili-

dad logarítmica del conjunto completo de datos \mathbf{O} más el logaritmo de la función de densidad de probabilidad previa $\log P_0(\lambda)$.

$$P(\lambda | \mathbf{O}) = P(\mathbf{O} | \lambda) + \log P_0(\lambda) \quad (3.64)$$

- *Aprendizaje diferencial (VAR)*: los métodos basados en el estudio de variaciones diferenciales proporcionan otro interesante punto en cuanto a los numerosos algoritmos de aprendizaje que tradicionalmente se han utilizado en problemas de tipo bayesiano [Mac97], [Att00]. En general, el problema de la inferencia probabilística que nos proponemos resolver mediante técnicas de aprendizaje diferencial, se reduce al cálculo de la distribución de probabilidad de los estados ocultos H condicionada a los valores observados E .

$$P(H | E) = \frac{P(H, E)}{P(E)} \quad (3.65)$$

Realmente, y debido a que no existen algoritmos que proporcionen una solución satisfactoria al problema de la inferencia o del aprendizaje, deberemos trabajar sobre una aproximación de la probabilidad condicional $P(H | E)$. Para ello se utilizará una familia de funciones de distribución de probabilidad condicional, $Q(H | E, \psi)$, donde ψ son el conjunto de parámetros variables que definen la distribución. De todas ellas se selecciona aquella distribución que minimice la divergencia de Kullback-Leibler (KL) $D(Q||P)$, respecto del conjunto de parámetros anteriormente mencionados. El desarrollo de toda esta teoría se puede aplicar al problema de entrenamiento de los HMM [JGS97] simplemente de tomar como variables independientes los parámetros que definen el HMM $\psi = \{\psi_1, \dots, \psi_N\}$ y agrupando los estados ocultos en el conjunto $S = \{S_0, \dots, S_T\}$. Definiendo las funciones de distribución adecuadas Q para cada uno de los parámetros del modelo, $Q(S)$, $Q(\pi)$, $Q(A)$, $Q(b)$ y minimizando su divergencia KL, se obtienen las ecuaciones correspondientes a los parámetros del HMM entrenado según las secuencias observadas. El problema radica en la selección de la familia de distribuciones de probabilidad condicionales Q , resultando finalmente que se pueden utilizar las funciones de distribución previas definidas para el algoritmo MAP.

Finalmente comentar que el caso del algoritmo EM no es más que una particularización del caso general de aprendizaje diferencial en el que se restringe la distribución probabilística estimada a un punto (δ de Dirac) [BC03].

- **Algoritmos genéticos.** Los algoritmos genéticos (genetic algorithms, GA) son un método de búsqueda estocástico capaz de optimizar el espacio de búsqueda de forma global (no local). En [KCMT01] se propone un nuevo método de entrenamiento basado en GA y en el algoritmo de Baum-Welch que resultan en un HMM con un número de estados óptimo. Utilizando su capacidad de búsqueda global y aplicando el GA al entrenamiento de los HMM, no sólo se puede encontrar el número de estados óptimo para el modelo, sino que también se pueden optimizar, en un solo paso, el resto de parámetros que lo definen. En resumen, el GA consta de los siguientes pasos: (i) el *mecanismo de codificación*, (ii) el *mecanismo de evaluación por ajuste*, (iii) el *mecanismo de selección* y (iv) el *mecanismo de reemplazo*. En la etapa en la que actúa el mecanismo de codificación, cada uno de los cromosomas del GA se asimilaría a un HMM, mientras que cada uno de los genes del cromosoma representaría cada uno de los parámetros del HMM asociado a su cromosoma. En la etapa de evaluación por ajuste se utiliza como criterio para el ajuste la función de probabilidad $P(\mathbf{O} | \lambda)$ referida en la *Sección 3.4.2*. Para el mecanismo de selección se utiliza la técnica de la ruleta. Únicamente operaciones genéticas, principalmente las mutaciones entre los estados, son las que permiten cambiar el número de estados de los HMMs asociados a los cromosomas, tratando de ajustarlos al número de estados óptimo. Finalmente se utiliza la reproducción regulada de los estados como estrategia de reemplazo. La principal ventaja del método propuesto [KCMT01] es que se encuentra la topología óptima del HMM en cualquier caso.
- **Simulated Annealing.** El algoritmo de *simulated annealing*¹ se basa en una aproximación heurística a la optimización probabilística. La idea básica es la del intercambio heurístico en el que, durante cada iteración, se transita desde una solución posible a otra más probable y situada

¹No existe una traducción que se ajuste a lo que el autor trata de expresar con *annealing*. Éste es un término metalúrgico que significa templar en una fragua, aquietar, calmar y que describe de forma certera el comportamiento matemático del algoritmo.

en la vecindad de la original. El algoritmo se detiene cuando se llega a un óptimo local. Como normalmente son varios los óptimos locales, se suele ejecutar el algoritmo varias veces, inicializándolo aleatoriamente desde puntos diferentes. Es curioso resaltar que, en esta técnica de entrenamiento, se permite que un resultado empeore siempre que este hecho permita liberar al algoritmo que se encuentra atrapado en un óptimo local [HAA96].

Variaciones sobre los HMM

En la literatura existen numerosos ejemplos que describen cómo, partiendo de la idea y definición originales de la teoría de los HMM, éstos han evolucionado en la búsqueda de mejoras que aumentarían sus prestaciones. Muchas de estas variaciones sobre los HMM están basadas en la generalización de modelos probabilísticos de tipo gráfico. A continuación se muestran algunos ejemplos.

- **HMM factorizados.** En los HMM factorizados (Factorial HMM, FHMM) los estados del modelo original se factorizan en un número variable. En este caso, el modelo tiene una topología distribuida [GJ97]. En este caso, la probabilidad conjunta para la secuencia de estados y de observaciones puede ser factorizada en diferentes HMM de la siguiente forma:

$$P(\{S_t, O_t\}) = P(S_1)P(O_1 | S_1) \prod_{t=2}^T P(S_t | S_{t-1})P(O_t | S_t) \quad (3.66)$$

La representación de cada uno de los estados del HMM original se puede entender como una colección de variables de estado donde cada una de ellas puede tomar M valores.

$$S_t = S_t^{(1)}, \dots, S_t^{(m)}, \dots, S_t^{(M)} \quad (3.67)$$

A este nuevo modelo nos referimos como FHMM, donde el espacio de estados está formado por el producto cruzado de todas las variables de estado en que se ha factorizado cada uno de los estados del HMM original. Si suponemos independencia entre los estados, obtenemos la si-

guiente función, que describe la probabilidad del estado:

$$P(S_t, S_{t-1}) = \prod_{m=1}^M P(S_t^{(m)} | S_{t-1}^{(m)}) \quad (3.68)$$

La comparación entre el HMM clásico y el modelo factorizado se aprecia en la *Figura 3.13*.

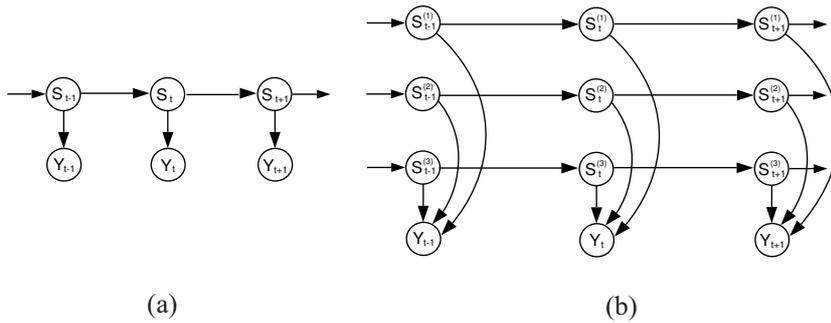


Figura 3.13: Diferencias entre un HMM (a) y un FHMM (b) donde, en el segundo caso, se aprecia la factorización $\times 3$ de los estados originales.

Debido a la naturaleza combinatorial de los estados ocultos, su resolución mediante el algoritmo de Baum-Welch es inabordable, por lo que el problema de inferencia se resuelve utilizando otros métodos (muestreo de Gibbs o ecuaciones de diferencias) [GJ97].

- HMM de entrada-salida.** En algunas aplicaciones resulta conveniente el separar la secuencia de observaciones en una secuencia que entra en el modelo y otra secuencia que se genera a su salida. En este caso el objetivo sería el de modelar la función de densidad de probabilidad de la secuencia de salida, dada la secuencia de entrada. Esto es lo que se conoce como HMM de entrada-salida (Input-Output HMM, IOHMM) cuya representación gráfica se puede apreciar en la *Figura 3.14* [BF95], [MJ96]. Dada una secuencia de vectores de entrada $\{X_t\}$, el modelo de probabilidad para un IOHMM responde a la ecuación:

$$P(\{S_t, O_t\} | \{X_t\}) = P(S_1 | X_1)P(O_1 | S_1, X_1) \prod_{t=2}^T P(S_t | S_{t-1}, X_t)P(O_t | S_t, X_t) \quad (3.69)$$

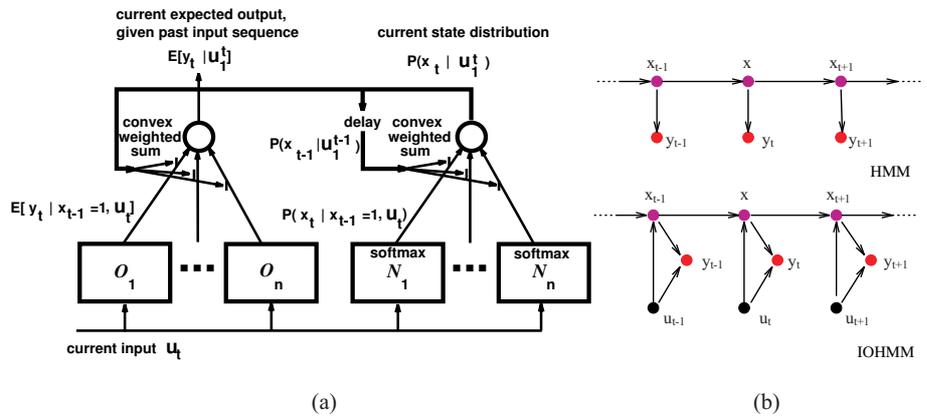


Figura 3.14: Arquitectura de un HMM de entrada-salida. (a) Arquitectura IOHMM propuesta. (b) inferior: Representación bayesiana de un IOHMM, superior: Representación bayesiana de un HMM estándar.

- **Árboles de decisión de Markov.** Los árboles de decisión de Markov (HMM decision trees, HMMDT) no es más que una generalización de los HMM [JGS97]. Su arquitectura podría entenderse como la de un HMM en el cual la variable que supone el número de estados existentes en cada instante de tiempo está factorizada (Figura 3.15). Desde otro punto de vista, estos árboles de decisión podrían entenderse como una serie de modelos de Markov dinámicos que van enlazando en cada instante de tiempo las variables de decisión [JJ94].
- **HMM difusos.** Como ya sabemos, en un HMM convencional los modelos son atomizados mediante el algoritmo de Baum-Welch. En este caso, cada una de las secuencias de observaciones es generada por la secuencia de estados de un modelo en concreto y sólo por uno. En el caso de HMM difusos, la secuencia observada en un instante de tiempo puede estar generada por más de un modelo a la vez. Este tipo de esquemas difusos son utilizados en [TW99] o [TK99] en el clustering (mediante un algoritmo k-medias *difuso*) de secuencias.
- **Los HMM y las Redes Neuronales.** Con la intención de superar las limitaciones que presentan los HMM tradicionales [BMR91], se propone la combinación entre éstos y las redes neuronales (Neural Networks, NNs)

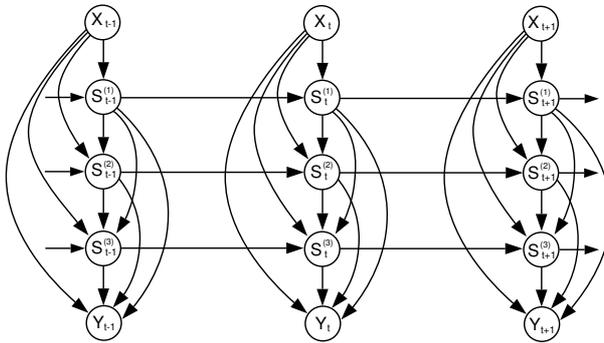


Figura 3.15: Árboles de decisión de Markov.

para conseguir un modelo híbrido que proporcione las ventajas de ambos combinando la potencia de cálculo de las redes neuronales con la característica secuencial de los HMM. El método consiste en utilizar las NNs para conseguir un aprendizaje "discriminante" que refine los clusters al mismo tiempo que optimice los modelos. La idea principal propone utilizar una multicapa de perceptrones (Multi-Layer Perceptron, MLP) para calcular las probabilidades de emisión de los HMM [BM90], [BC96]. Si se asocia la salida de una MLP a un estado concreto del modelo q_k , es posible entrenar la MLP para estimar las probabilidades de emisión (con una mayor precisión y capacidad discriminante, y sin utilizar ningún tipo de hipótesis al respecto de la distribución probabilísticas de los datos de entrada) de la forma $p(q_k | o_n)$ donde o_n (vector de características), se utiliza como entrada (Figura 3.16). Nos referiremos a la probabilidad $p(q_k | o_n)$ como probabilidad de Bayes o probabilidad *a posteriori* que podemos transformar en un valor de probabilidad susceptible de ser utilizadas como probabilidad de emisión de un estado, utilizando el teorema de Bayes:

$$p(o_n | q_k) = \frac{p(q_k | o_n)p(o_n)}{p(q_k)} \quad (3.70)$$

- **HMM auto-regresivo.** En la Figura 3.17 se muestra la estructura de un HMM auto-regresivo (Auto-Regressive HMM, ARHMM) que se utiliza para reducir el efecto "embudo" generado por los estados S_t , interconec-

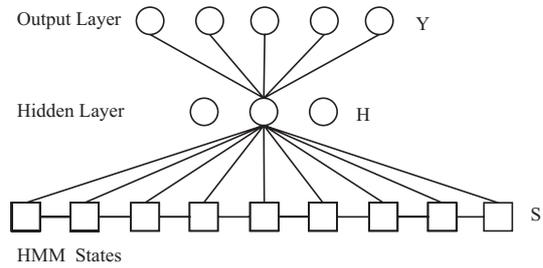


Figura 3.16: Cada uno de los estados S del HMM aparece conectado con todos los componentes la capa oculta común H y, a su vez, cada un de los componentes de esta capa oculta está conectado con todas y cada una de las salidas Y , mediante las que se calcula la probabilidad de emisión.

tando para ello las funciones de probabilidad de emisión para permitir que Y_{t-1} participe en la predicción realizada por Y_t , con lo que se consiguen modelos con una mayor probabilidad [Mur02]. Si Y resulta ser continua, podemos definir su función de densidad de probabilidad como:

$$P(Y_t = y_t \mid S_t = i, Y_{t-1} = y_{t-1}) = N(y_t; R_i y_{t-1}, C_i) \quad (3.71)$$

donde R_i es la matriz de regresión dado S_t en el estado i .

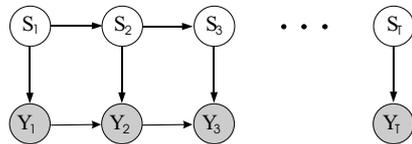


Figura 3.17: HMM auto-regresivo.

- **Modelos de Markov Enterrados.** Los modelos de Markov enterrados (Buried Markov Models, BMM) [Bi198], [Mur02] no son más que una generalización de los ARHMM presentados en el epígrafe anterior, pero en este caso la dependencia entre los nodos puede ser no-lineal *Figura 3.18*.

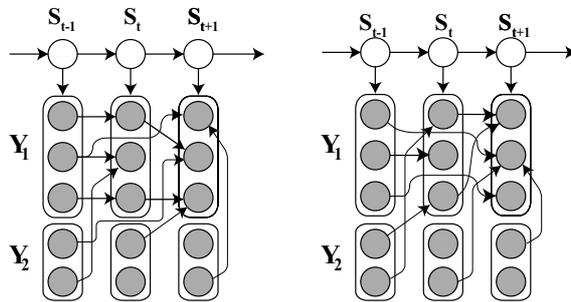


Figura 3.18: Modelos de Markov enterrados. Dependiendo del valor de las variables ocultas, S_t , la estructura y dependencia entre las componentes de las secuencias observadas, Y_t , varía.

- Modelos de Markov con memoria entrelazada.** Una de las aproximaciones más simples que se utilizan para modelar conjuntos de datos de tipo secuencial son los n -gramas. Los n -gramas no son más que cadenas de Markov de orden n (donde, por ejemplo, para $n = 2$ tenemos un bigrama, para $n = 3$ un trigramo, etc.). Cuando Z_t es una variable aleatoria discreta que puede tomar muchos valores diferentes (por ejemplo, representando las palabras de un diccionario), entonces puede que no se disponga de los datos suficientes para poder estimar adecuadamente $P(Z_t = k | Z_{t-1} = j, Z_{t-2} = i)$. Para resolver este problema aproximando de alguna manera el HMM original, se utiliza una mezcla de modelos de Markov de orden inferior:

$$\begin{aligned}
 P(Z_t | Z_{t-1}, Z_{t-2}) &= \alpha_3(Z_{t-1}, Z_{t-2})f(Z_t | Z_{t-1}, Z_{t-2}) \\
 &+ \alpha_2(Z_{t-1}, Z_{t-2})f(Z_t | Z_{t-1}) + \alpha_1(Z_{t-1}, Z_{t-2})f(Z_t)
 \end{aligned}
 \tag{3.72}$$

donde los coeficientes α pueden depender opcionalmente de valores anteriores y donde $f(\cdot)$ es una distribución de probabilidad condicional [Jel99].

- HMMs acoplados.** En un HMM acoplado (coupled HMM, CHMM) [SJ95], [Bra96] and [RPR00] las variables ocultas interactúan localmente con

sus vecinas. Se trata de una nueva interpretación de los sistemas multi-dimensionales que ya han sido introducidos cuando anteriormente se presentaban los HMMNN y los IOHMM. El estado actual del modelo depende de los estados de su propia cadena S_t^i y de los resultados obtenidos por la cadena vecina Y_{t-1}^{i-1} en un instante de tiempo anterior 3.19.

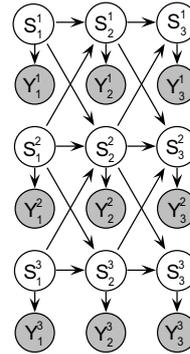


Figura 3.19: HMM acoplado de 3 cadenas.

- **HMM jerárquico.** Los HMM jerárquicos (Hierarchical HMM, HHMM) [FST98] suponen una extensión de los HMM diseñados específicamente para modelar estructuras de tipo jerárquico. En un HHMM, los estados del autómata pueden emitir bien observaciones únicas, bien cadenas de observaciones. Aquellos estados que emiten observaciones únicas se llaman *productivos*, y los que generan cadenas son los estados *abstractos*. Las cadenas emitidas desde los estados abstractos están a su vez modeladas por sub-HMMs, que pueden ser utilizados recursivamente. Cuando el sub-HMM genera su observación correspondiente, automáticamente devuelve el control al estado desde el que se le había invocado. Para memorizar la cadena de estados se utiliza una pila de memoria.
- **HMM de duración variable (semi-Markov).** En la expresión gráfica de un HMM, un arco que parte de un estado para llegar al mismo estado se utiliza para modelar secuencias de duración variable (espera variable en el mismo estado). En concreto, el problema que se plantea es el de permanecer en el estado i durante d pasos: $p(d) = (1 - p)p^{d-1}$ donde

$p = A(i, i)$ es la probabilidad de permanencia en el mismo estado. Para modelar duraciones variables de una forma más general, se utilizan los modelos de semi-Markov [Mur02], llamados de esta forma porque para predecir el estado siguiente se necesita conocer no sólo el resultado del estado anterior, sino también el tiempo de permanencia en el estado actual. Gráficamente se pueden representar este tipo de modelos tal y como se muestra en la *Figura 3.20*. La razón por la que no aparecen conexiones desde el estado S al F es porque el proceso que decide finalizar la permanencia en el estado S_i es un proceso determinista. En concreto, el nivel inferior es el encargado de decidir durante cuanto tiempo se permanece en determinado estado; cuando su contador llega a 0, el nodo F se activa, el nodo original S puede entonces cambiar de estado, y el contador, S^D se borra.

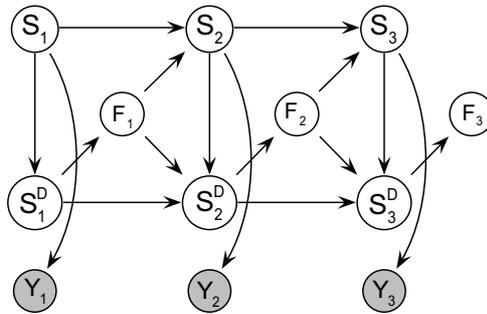


Figura 3.20: HMM de duración variable modelado como un HMM de dos niveles donde S_t representa el estado, y S_t^D determina en tiempo de permanencia en ese estado.

- Modelos segmentados.** La idea básica de estos modelos [ODK96] es que cada estado del HMM original genere toda una secuencia de observaciones en vez de producir una única observación. La diferencia entre un HMM segmentado (segmented HMM, SHMM) y un HMM tradicional reside en que, en los HMM segmentados, la longitud de la secuencia generada desde el estado S_i se determina a través de una variable aleatoria adicional l_i (*Figura 3.21*). De esta manera, es la variable l_i la que de-

termina el número de segmentos a generar desde cada estado.

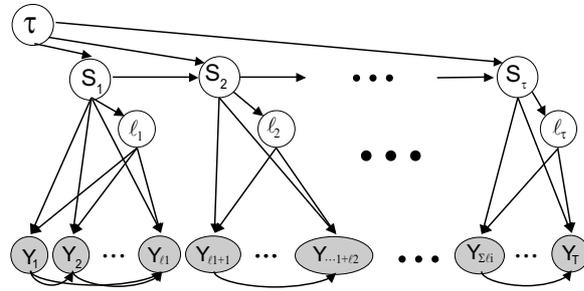


Figura 3.21: esquema de una HMM segmentado [Mur02]. Los nodos Y_t son observables y el resto de nodos permanecen ocultos.

- **Otras variantes y mejoras.** A continuación se enumeran toda una serie de métodos y técnicas que, basadas en los HMM, se han desarrollado de una forma más teórica que práctica.
 - En [BC03] se proponen los modelos de HMM con un número infinito de estados ocultos y donde se utiliza para resolverlo el teorema de Dirichlet.
 - En [TLF99] y [Sco02] se introducen los HMM Monte-Carlo (MCHMM) donde se utilizan árboles de funciones de densidad de probabilidad no parametrizadas y se utilizan para modelar sistemas continuos (los espacios del conjunto de observaciones y del conjunto de estados son continuos).
 - En [Ber99] se extiende el orden de dependencias a dos (Modelo de Markov de Doble Cadena, DCMM), combinando dos cadenas de Markov: una (la de observaciones) no homogénea y otra (la de estados) homogénea.
 - En [Laf95] se desarrollan los *Modelos de Markov-Gibbs*, donde se utilizan las distribuciones de Gibbs para modelar las probabilidades de transición entre los estados, y donde la estimación y optimización de los parámetros del modelo se realiza con el algoritmo EM.

- Una nueva variante de los HMM discretos se propone en [CZO98]. En este caso, se utiliza la entropía de los parámetros que caracterizan los estados del modelo para ponderar la probabilidad de emisión. Por otra parte en [SW96], también se proponen HMMs basados en la función de entropía calculada entre los parámetros del HMM calculado en la iteración t y los del HMM calculado en la iteración $t - 1$. Finalmente en [AH98] también se propone un nuevo método de entrenamiento.
- Finalmente Tran [TW00] presenta los llamados *HMM de niveles de tramas* (*Frame-Level HMM*) donde la dependencia entre estados se basa en la probabilidad de cada observación en particular, en contraste con la visión clásica de los HMM donde la dependencia entre los estados viene dada por la probabilidad de la secuencia de observaciones.

Aplicaciones de los HMM

Ni la teoría de los HMM ni sus aplicaciones en el mundo de la ingeniería son nuevas. Sus bases fueron publicadas por Baum [BP66] y sus colaboradores en una serie de artículos entre finales de la década de los sesenta y principios de los setenta, y sus primeras aplicaciones en el campo del procesamiento del habla fueron implementadas por Baker [Bak75] y posteriormente Jelinek [Jel99] en IBM. Por otra parte los HMM tienen otros muchos campos de aplicación: biología molecular, robótica, reconocimiento de patrones de escritura, etc. A continuación se reseñan las principales aplicaciones que, con estos modelos, se han desarrollado en las diferentes áreas:

- **Procesamiento del habla.** Ha sido desde siempre el campo de aplicación tradicional de la teoría de HMM. Las primeras aplicaciones en este sentido fueron propuestas de forma independiente por Baker [Bak75] y por el grupo de investigación de IBM [Jel76]. Desde entonces se han publicado un gran número de artículos que desarrollan diversas implementaciones de los HMM, entre los que cabe citar, por ejemplo a [Gal00], [BC96], [BM90], [TW99], [BF95], [GJ97].

- **Biología molecular.** En este caso los HMM se aplican para resolver problemas de modelado estadístico, búsquedas en bases de datos, alineamiento de secuencias proteínicas pertenecientes a una misma familia, [Ke94], [BB00], etc.
- **Sistemas dinámicos y robótica.** En [Cha94] se utilizan HMMs para la conmutación lineal de modelos. En [Smy94] se aplican para la detección y tratamiento de errores en sistemas lineales. El grupo de investigación de Yang [YXC97] o [MYD96] centran su interés en el aprendizaje y simulación robotizada de habilidades humanas incluido el reconocimiento de gestos.
- **Reconocimiento de textos.** Se trata de un campo de la disciplina de visión por computador (VxC) con aplicación en múltiples áreas. Como ejemplo se proponen [AB95], [VK92].
- **Economía.** La predicción de la evolución de los indicadores en el mercado de valores es un tópico en este campo. Weigend [WS98] introdujo a expertos en HMM para realizar predicciones de bolsa. Otro ejemplo de aplicación de los HMM a los mercados bursátiles se puede encontrar en [MZ00].
- **Análisis de imágenes.** Un campo de aplicación donde los HMM han demostrado ser de utilidad es en el procesamiento de imágenes reales bidimensionales. Un artículo de gran impacto en el ámbito del procesamiento y análisis de imágenes y que proponía un modelo de campos de Markov aleatorio (Markov Random Field, MRF) fue propuesto en 1984 por [GG84]. El mismo problema se puede resolver aplicando cadenas de Markov [AEH99]. Los HMM también se han incorporado a la teoría de la transformada wavelet [Cro98], [Nov02].
- **Procesamiento de señales biomédicas.** En [Obe01] los HMM se proponen para identificar señales electroencefalográficas (EEG) mientras el paciente imagina el movimiento de su mano derecha o de su mano izquierda. Otra de sus múltiples aplicaciones se produce en el análisis de señales biomédicas multiregistro [RB94] o [PR98] donde los HMM se utilizan para detectar cambios en niveles de continua, correlación entre señales o cambios de frecuencia. En el área de la *Inteligencia Artificial* [PMM02]

se propone utilizar HMM para la segmentación de tareas cognitivas. Por otra parte, [RMR02] sugiere utilizarlos para la detección y análisis de las etapas del sueño en señales EEG. El mismo autor, [RR02] implementa HMM para analizar pequeños intervalos cardíacos (en ECGs) y otros registros de la actividad respiratoria. Por supuesto también se han utilizado para el procesamiento de señales ECG [LBBC01] y [CSC90] tanto para la detección del complejo *QRS* como para la identificación de la onda *P*. En [CC98], se propone una aproximación mediante HMMs para la clasificación ECG. Finalmente Koski [KJ96] realiza un modelado de la señal ECG mediante HMM.

3.5. Reconocimiento de patrones

A través del reconocimiento de formas [DGL96] [DHS01] [Fu82] [TK99] se estudia la construcción de sistemas que, de forma automática, sean capaces de emular aspectos perceptivos propios del comportamiento humano (por lo general relacionados con el habla o con la visión).

El paradigma clásico que sustenta la disciplina del reconocimiento de patrones o formas es el de la clasificación de los objetos a analizar. De acuerdo con esto, un sistema de reconocimiento de formas puede entenderse como una *caja negra* a cuya entrada tendríamos la señal obtenida mediante una serie de transductores que miden el objeto a reconocer (en nuestro caso la señal electrocardiográfica, medida con el Holter). A la salida del clasificador obtendríamos una etiqueta de clase (por latido) perteneciente a un conjunto finito de etiquetas. Nuestro objetivo es el de reconocer el objeto minimizando la probabilidad de error en la clasificación o, más generalmente, el riesgo total.

Las principales aproximaciones al reconocimiento de formas son: *la geométrica o estadística* [DGL96] [DHS01] [TK99]; y *la sintáctica o estructural* [Fu82] [TK99]. La segunda se basa en la teoría de lenguajes formales mientras que la primera, posiblemente la más extendida, se basa en la teoría estadística de la decisión. Como ya hemos comentado anteriormente en nuestro caso centraremos la atención en los métodos de aproximación estadística. En este tipo de aproximación podemos dividir los sistemas clasificadores en *paramétricos* y *no paramétricos* [Jua99]. Los clasificadores paramétricos asumen que las densidades

de probabilidad condicionales de las clases (distribución estadística de las muestras de la clase en el correspondiente espacio muestral) tienen una forma conocida, habitualmente (una mezcla de) gaussiana(s), que depende de una serie de parámetros desconocidos. En este caso, pues, el aprendizaje se reduce a aplicar técnicas convencionales de la estadística para estimar los parámetros desconocidos. Por otra parte, los clasificadores no paramétricos no asumen que las densidades condicionales tengan formas conocidas. Algunas de estas técnicas tratan de estimar las densidades condicionales localmente, a partir de muestras de entrenamiento. Otras, optan por estimar directamente las probabilidades a posteriori. El objetivo último de este tipo de clasificadores será el de particionar el espacio de representación del mismo modo que el clasificador de Bayes, aproximándose en lo posible a las fronteras de separación óptimas entre clases.

En un ámbito más general, y según se disponga o no inicialmente de elementos convenientemente etiquetados, también podemos distinguir entre métodos de clasificación supervisados (que se calculan y entrenan a partir de las muestras etiquetadas) y métodos de clasificación no supervisados (o métodos de clustering).

3.5.1. Métodos supervisados

Para aplicar cualquiera de los métodos de clasificación supervisados (*supervised pattern recognition*) que existen en la literatura será necesario inicialmente contar con una serie de observaciones de entrenamiento y otras de test. Después de realizar los pasos previos correspondientes al preprocesamiento, y para empezar con la fase de agrupación de los latidos componentes de la señal, se pueden utilizar cualquiera de las técnicas descritas en el trabajo [CDA94a], algunas de las cuales se comentan a continuación. Para concluir con este apartado, en [CDA94b] se puede encontrar una revisión de todas aquellas aplicaciones que utilizan métodos de clasificación supervisados para agrupar señales biológicas. En esta revisión se detallan todos los elementos involucrados en el proceso: fuente y propiedades de cada señal, extracción y selección de características y método de clasificación.

- *Métodos estadísticos.* Cada observación se representa en términos de ca-

racterísticas, considerándola como un punto en un espacio d -dimensional. El objetivo es extraer aquellas propiedades intrínsecas que permitan a los vectores que pertenezcan a diferentes categorías ocupar regiones disjuntas en dicho espacio d -dimensional de características [JDM00] [SMLZ98].

- *Métodos basados en mínima distancia.* Clasifican un patrón basándose en su distancia a un prototipo de clase. Entre estos métodos destacan los basados en la distancia Euclídea y en la distancia de *Mahalanobis* [CDA94a] [MPC⁺81] [dCMR99].
- *Métodos sintácticos.* En muchos problemas complejos de reconocimiento de formas, es mejor adoptar una perspectiva jerárquica, donde cada patrón se considera compuesto por una serie de subpatrones simples, denominados primitivas, de manera que se establece una analogía formal entre la estructura del patrón complejo y la sintaxis de un lenguaje. La gramática de cada clase debe ser inferida a partir del conjunto de muestras de entrenamiento disponibles [JDM00] [KJM95] [TS90].
- *Redes neuronales.* Los métodos basados en redes neuronales se utilizan en la clasificación cuando se dispone de poca información a priori sobre las clases. Suelen estar estructurados en capas, en cuyo caso contienen una capa de entrada y otra de salida, pudiendo además incluir alguna capa intermedia, compuesta cada una de conexiones y nodos. Cada conexión estará ponderada según un coeficiente. Las redes neuronales se entrenan a partir de un conjunto de observaciones de entrenamiento, y éstas iterativamente ajustan los pesos de las conexiones para obtener la salida deseada. En [PD00] se utiliza un perceptrón multicapa para la clasificación de diferentes tipos de electroencefalogramas (EEG).

3.5.2. Métodos no supervisados

En muchas aplicaciones de reconocimiento de formas, puede ser muy difícil o costoso disponer de un conjunto de muestras de entrenamiento. Los métodos no supervisados (*unsupervised pattern recognition* o métodos de *clustering*) [Jai88] [JMF99] [Har74] [Jua99] presentan como objetivo el del agrupamiento de los objetos en clases naturales o clusters, a partir de observaciones sin etiquetar, y basándose en similitudes medidas entre los objetos a

clasificar. Al no tener a priori un conocimiento de las clases existentes, ni disponer de un conjunto de objetos etiquetados, resulta imposible un diseño de funciones discriminantes.

Existen numerosos algoritmos desarrollados para el clustering de datos para los que Berkhin propone en [Ber02] una esclarecedora clasificación. Dentro de las últimas tendencias relativas a los procesos de clustering en general, cabría destacar los trabajos referidos en [Spa80] [Kol] [HKT01] [Gho02]. Una buena introducción a las técnicas de clustering y data mining se encuentra en [HK01]. Por otro lado, la relación existente entre las técnicas de clustering y una gran cantidad de campos científicos (estadística, reconocimiento de patrones, reconocimiento del habla y de la escritura, visión por computador y segmentación de imágenes, estimaciones estadísticas de funciones de densidad, procesamiento de imágenes, cuantización vectorial, modelado de datos, etc.) queda reflejada y convenientemente documentada en [Ber02].

Formulación matemática del proceso de clustering

El clustering es un proceso natural inherente a la actividad del cerebro humano que siempre intenta ordenar y clasificar la información recibida como estímulo desde fuentes externas. En primer lugar, y para referirnos a los elementos del proceso de clustering de una manera formal, definimos:

- P : conjunto de n puntos en \mathbb{R}^d . Corresponde al conjunto de n objetos d -dimensionales $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ (en un espacio vectorial) o de carácter no necesariamente vectorial (en un espacio métrico o pseudométrico) a agrupar.
- $d(x, y)$: con $x, y \in P$. Es la función de disimilitud entre dos elementos del conjunto.
- $Q \subset P$: conjunto de k prototipos o centroides $\{q_1, q_2, \dots, q_k\}$, con $k \leq n$.
- $C = \{C_1, C_2, \dots, C_k\}$. Es la partición de P optimizada según la función criterio que se utilice en cada caso.

Para la formulación matemática del proceso y siendo el conjunto de datos P , definimos como un k -clustering de P a su clasificación en k particiones

(clusters), C_1, \dots, C_k , de tal forma que se cumplan las tres condiciones siguientes:

- $C_i \neq \emptyset, \quad i = 1, 2, \dots, k$
- $\bigcup_{i=1}^k C_i = P$
- $C_i \cap C_j = \emptyset, \quad i \neq j \quad i, j = 1, \dots, k$

Cabe destacar que, bajo las anteriores premisas, cada uno de los vectores (elementos) a clasificar sólo puede pertenecer a un único cluster, lo que se conoce como *hard* o *crisp* clustering. Una definición alternativa (en términos de clustering *difuso* donde se trata de clasificar los datos en k clusters caracterizados por k funciones miembro (u_j) donde, en este caso, cada uno de los vectores \mathcal{X} pertenece o puede pertenecer a más de un cluster de forma simultánea, propiedad que se cuantifica mediante el correspondiente valor de la función miembro u_j , definida en el intervalo $[0, 1]$.

Descripción del proceso de clustering

En [DHS01] y [TK99] se describen los pasos básicos que se deben seguir en todo proceso de clustering:

- **Extracción y reducción de características**, ambos procesos han quedado convenientemente descritos en la *Sección 3.2* y en la *Sección 3.4*.
- **Medida de disimilitud**, mediante la que se define la similitud (proximidad, medida de similitud o MS) o disimilitud (separación, medida de disimilitud o MD) entre pares de objetos, objetos y clusters y pares de clusters. Para calcularla correctamente será necesario asegurar que ninguna de las características seleccionadas domina sobre las otras ya que, en caso contrario, se podría estar enmascarando información que no quedaría reflejada en la medida (distorsionando los posteriores resultados del clustering). Tanto los objetos cuyas características responden a una distribución continua (como sería el caso de la caracterización de la amplitud y duración de los latidos de un ECG, ver *Capítulo 4*), como los centroides de los clusters definidos por éstos, se pueden considerar

como puntos de un espacio multidimensional (la dimensión quedaría definida por el número de características del objeto). En este caso la medida de similitud o de disimilitud se podría calcular a través de distancias como la euclídea, la distancia de Manhattan, la métrica de Mahalanobis [TK99] o incluso utilizando la correlación entre los objetos (distancia de Tanimoto [Tan58]).

Por otra parte, y si queremos calcular la distancia entre dos clusters en términos de funciones de densidad de probabilidad (*probability density function*, pdf) que los caracterizan, podremos utilizar la distancia de Bhattacharya. Dada la pdf del cluster, también es posible obtener una medida de disimilitud entre éste y el objeto a clasificar en términos de probabilidad [CS96].

Las propiedades que debe cumplir una medida de disimilitud d , calculada entre dos objetos cualesquiera x_i, x_j de un conjunto con n elementos son [Cop68]:

- $d(x_i, x_j) \geq 0, \quad \forall i, j \mid i \neq j, \quad i, j = 1 \dots n$
- $d(x_i, x_i) = 0, \quad \forall i \mid i = 1 \dots n$
- $d(x_i, x_j) = d(x_j, x_i), \quad \forall i, j \mid i \neq j, \quad i, j = 1 \dots n$
- $d(x_i, x_k) \leq d(x_i, x_j) + d(x_j, x_k)$
 $\forall i, j, k \mid i \neq j \neq k, \quad i, j, k = 1 \dots n$

Algunas medidas de disimilitud útiles en la práctica no cumplen todas estas propiedades, y entonces se denominan semisimétricas, pseudosimétricas, etc [Cop68].

Si la métrica utilizada mide la semejanza entre los objetos del conjunto anterior, diremos que nos encontramos ante una función de similitud s , cuyas características son [TK99]:

- $s(x_i, x_j) \geq 0, \quad \forall i, j \mid i \neq j, \quad i, j = 1 \dots n$
- $s(x_i, x_i) = 0, \quad \forall i \mid i = 1 \dots n$
- $s(x_i, x_j) = s(x_j, x_i), \quad \forall i, j \mid i \neq j, \quad i, j = 1 \dots n$
- $s(x_i, x_j) \cdot s(x_j, x_k) \leq [s(x_i, x_j) + s(x_j, x_k)] \cdot s(x_i, x_k)$
 $\forall i, j, k \mid i \neq j \neq k, \quad i, j, k = 1 \dots n$

- **Funciones criterio**, son utilizadas para evaluar y comparar la calidad de las diferentes particiones de los clusters y determinar cuál de ellas representa y agrupa la información más correctamente. En el *clustering paramétrico*, la calidad de una partición se determina en términos probabilísticos al comparar la partición con los datos iniciales [CS96] utilizando para ello dos factores: por un lado la probabilidad inicial del modelo y por otro lado la probabilidad de los datos calculada sobre ese modelo. Mediante la probabilidad inicial se favorece a todos aquellos modelos con pocas particiones, mientras que si ponderamos la probabilidad de los datos sobre el modelo, estaremos favoreciendo los modelos de tamaño grande y con un número elevado de clusters. La mejor de las particiones seleccionada siguiendo el criterio presentado será aquella que presente un equilibrio entre los dos factores anteriores. En el *clustering no paramétrico*, la calidad de la partición se evalúa en términos de su capacidad para maximizar la disimilitud intra-cluster (diferencia entre los distintos clusters) minimizando al mismo tiempo la disimilitud inter-cluster (diferencias entre los objetos dentro de un cluster). Funciones criterio como la del error cuadrático medio (Mean Squared Error, MSE) [DHS01], estiman la calidad de la partición en términos de error o coherencia (media de las desviaciones de los datos sobre el centroide de cada uno de los clusters existentes).

- **Algoritmos de clustering**, que revelan la estructura del conjunto de datos. Una vez elegidas la medida de disimilitud y la función criterio a emplear, se debe seleccionar el algoritmo de clustering. Existen muchos métodos para llevar a cabo este proceso. Dependiendo del método empleado se pueden llegar a obtener distintos resultados aunque los datos de partida y la medida de disimilitud sean los mismos [JDM00]. Una solución general exigiría analizar todas las posibles asignaciones de los n objetos en k clusters. Sin embargo esto no se utiliza en la práctica debido a su elevado coste computacional, que exigiría analizar $\binom{n}{k}$ particiones distintas. Por lo tanto, y para seleccionar el algoritmo de clustering a utilizar, deberemos tener en cuenta su coste computacional y, además, la necesidad de memoria del sistema (representada por la cantidad de medidas a almacenar) y que en algunos casos hace inviable el clustering en conjuntos con una gran cantidad de datos. Existen, sin embargo, tres

técnicas básicas de agrupamiento en las que se basan la mayoría de los métodos y que se detallan a continuación:

- *Algoritmos de agrupamiento secuencial*, en los que las características de los objetos a clasificar se procesan un número limitado de veces. El resultado final del agrupamiento depende enormemente del número de iteraciones realizadas al procesar los vectores de características [Hal67].
- *Algoritmos de agrupamiento particional*, clasificados en la categoría de algoritmos no jerárquicos y no paramétricos, cuyo objetivo es el de minimizar la variabilidad intraclase maximizando la variabilidad interclase (Figura 3.22) [JDM00]. En términos generales, podemos decir que los métodos no paramétricos tratan de encontrar regiones de alta densidad de datos separadas mediante regiones de baja densidad. Para ello se descompone directamente el conjunto de datos en un conjunto de clusters disjuntos, minimizando una función criterio que enfatiza su estructura local, asignando clusters a máximos locales en función de la densidad de probabilidad, o en la estructura global. Es el tipo usado con más frecuencia en el reconocimiento de formas. Dependiendo de si el objeto a clasificar se asigna a uno o a varios clusters nos encontramos con dos tipos de algoritmos particionales, los de clasificación dura (*hard partitioning*) y los de clasificación blanda (o *soft partitioning*). Algunos ejemplos de este tipo de algoritmos los constituyen el modelo ISODATA [DHS01], el *K*-Medias (*K*-Means) [TK99], el modelo AUTOCLASS [SC94] y los presentados en [KK96] y [Ati90].
- *Algoritmos de agrupamiento jerárquico*, que proporcionan como resultado una secuencia anidada de grupos, representada en forma de dendograma o árbol (Figura 3.23) [JDM00]. El clustering jerárquico se puede llevar a cabo de dos modos distintos, bien uniendo clusters pequeños para formar otros mayores (clustering aglomerativo) o, por el contrario, dividiendo clusters grandes en otros menores (clustering divisivo). El resultado es un árbol de clusters denominado dendograma, el cual muestra la forma en la que están relacionados dichos clusters. Mediante el corte del dendograma en un

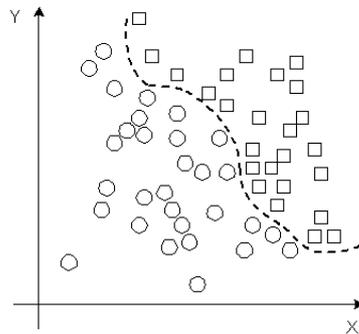


Figura 3.22: Ejemplo de clustering parcial en el que se trata de maximizar la varianza intercluster y minimizar la varianza intracluster. En este caso se han utilizado dos clusters C_1 y C_2 .

cierto nivel se obtiene el clustering de los datos en grupos disjuntos. Ejemplos de clustering aglomerativo los podemos encontrar en [TK99], mientras que, citando algunos de los métodos de clustering divisivo aparecen los de COBWEB [FXZ92] o ITERATE [BWF98].

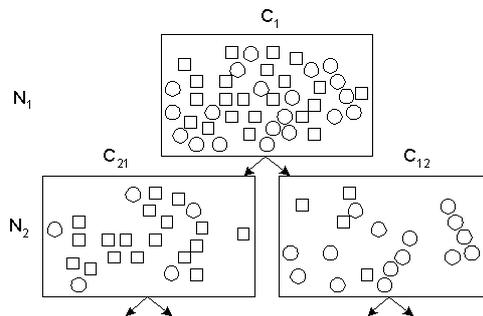


Figura 3.23: Ejemplo de clustering jerárquico. En este caso se va dividiendo el cluster inmediatamente anterior hasta conseguir una estructura con los diferentes grupos divididos según la función criterio que se trata de minimizar. Partiendo de un conjunto de objetos a agrupar C_1 , se obtienen dos subconjuntos C_{11} y C_{12} , en el nivel de descomposición N_2 . En caso de considerarse oportuno, estos subconjuntos se podrían dividir a su vez hasta conseguir unas particiones adecuadas según la función criterio.

Además, existen otros algoritmos de clustering no recogidos en las anteriores categorías, donde incluiríamos los siguientes:

- *Métodos de rama y frontera (branch and bound)*. Nos proporcionan un óptimo global sin la necesidad de tener que analizar todos los posibles clusters [KNF95].
 - *Algoritmos genéticos*. Utilizan un número de clusters inicial para, a partir de él, refinarlo y mejorarlo de acuerdo a un criterio determinado [Sch97].
 - *Métodos de relajación estocásticos*. Son capaces de garantizar, bajo ciertas condiciones, la convergencia hacia un óptimo global de los clusters [BH92], [Hof97].
 - *Algoritmos de aprendizaje competitivo*. Compuestos por una serie de iteraciones que no utilizan ningún tipo de función de coste. A partir de las iteraciones se generan varios clusters convergiendo hacia uno de ellos, de acuerdo con una matriz de distancias [RZ86], [Koh90].
 - *Algoritmos basados en transformaciones morfológicas*. Mediante las que se consiguen una separación mucho más clara de los clusters involucrados [TK99].
 - *Algoritmos de búsqueda del valle (valley-seeking)*. Operan con los vectores de características como si fueran parte de una variable aleatoria multidimensional [Fuk90].
- **Inicialización de los clusters**. Una correcta elección en el número de clusters k puede resultar crítica. Diferentes valores de k pueden dar como resultado diferentes tipos de clusters. Pero además, la inicialización de los clusters también resulta esencial ya que algunos de ellos pueden quedar vacíos si el centroide inicial queda lejos de la distribución de los datos. Por otra parte, también será necesario el agrupamiento inicial de los datos. Existen muchas formas de realizar dicho agrupamiento, entendiéndolo como la selección de los centroides iniciales o la confección de una partición inicial que permita llegar a la agrupación final con un menor coste computacional [Jua99]. En este caso, el método más sencillo probablemente sea el de inicialización aleatoria [Jua99], en el cual

se seleccionan los centroides aleatoriamente de entre el conjunto de objetos a agrupar, aunque también se pueden tomar como centroides los primeros k objetos consecutivos del conjunto. Es evidente que esta técnica es la que menor coste computacional presenta, sin embargo puede dar lugar a una clasificación con una baja variabilidad intercluster si los objetos se encuentran muy cercanos entre sí desde el punto de vista de la medida de disimilitud empleada [Jua99].

- *Validación de los resultados*, cuando, una vez obtenidos los resultados del clustering, debemos verificar de alguna manera su verosimilitud. Para ello se diseñan un conjunto de pruebas apropiadas (comparación con objetos previamente etiquetados, inspección visual, etc.).
- *Interpretación de los resultados*, ya que, en muchas ocasiones, el experto en el campo de aplicación en el que se ha usado el clustering tiene que integrar los resultados obtenidos con los de otros experimentos para poder obtener así las conclusiones globales del experimento.

Clustering de series temporales de datos

El caso del agrupamiento de los latidos de un ECG, y que es el motivo del presente trabajo, podría resumirse en el problema del agrupamiento de series de datos temporales que se plantea en este apartado. A diferencia de las series de datos estáticas, que se utilizan para describir las propiedades de los sistemas invariantes en el tiempo, las series de datos temporales sí que incluyen todos aquellos aspectos del comportamiento dinámico de los sistemas que describen. Cada una de las características de una serie de datos temporal tiene que ser forzosamente bidimensional ya que se incluye al tiempo implícitamente como segunda dimensión. En una serie de datos descrita por K características y donde cada característica está representada por una secuencia de T valores temporales, un objeto de la serie x_i , puede representarse mediante una matriz $K \times T$ de la forma:

$$x_i = \begin{pmatrix} F_{11}^i & F_{12}^i & \cdots & F_{1T}^i \\ F_{21}^i & F_{22}^i & \cdots & F_{2T}^i \\ \vdots & \vdots & \vdots & \vdots \\ F_{K1}^i & F_{K2}^i & \cdots & F_{KT}^i \end{pmatrix} \quad (3.73)$$

donde F_{kt}^i representa el valor de la característica temporal k en el instante de tiempo t para el objeto x_i .

Una manera muy simplista de interpretar una serie de datos temporal sería la de suponer que las características del objeto en un instante de tiempo determinado son independientes entre sí. Lo que significa que, para un objeto de la serie x_i , se podría descomponer en T objetos estáticos diferentes de la forma $x_1^i, x_2^i, \dots, x_T^i$, donde cada uno de ellos estaría definido por un vector de características de tamaño K tal y como sigue:

$$\begin{aligned} x_1^i &= [F_{11}^i, F_{21}^i, \dots, F_{K1}^i] \\ x_2^i &= [F_{12}^i, F_{22}^i, \dots, F_{K2}^i] \\ &\vdots \\ x_T^i &= [F_{1T}^i, F_{2T}^i, \dots, F_{KT}^i] \end{aligned} \quad (3.74)$$

De esta manera obtenemos un nuevo conjunto de datos de dimensión $L \times T$ (donde L es el tamaño del conjunto de datos temporales), reduciendo la secuencia temporal de los datos a su representación estática. En este punto es cuando podríamos aplicar sobre los datos todos aquellos algoritmos de clustering descritos en la sección anterior. Para el clustering de latidos de un ECG Holter, y una vez segmentado en sus n latidos, se podría considerar como serie de datos temporales a la señal constituida por un único latido cuyas K características (variables aleatorias del proceso estocástico, considerado estacionario para un desplazamiento t_0 igual al período de los latidos y cuyas realizaciones son los propios latidos de la señal ECG) evolucionan con el tiempo (sería como superponer temporalmente los latidos sucesivos como si se tratara de una película). En este caso $L = 1$ y $T = n$ donde el número de características del latido se define en el proceso de extracción.

A continuación se presentan tres de los métodos de clustering utilizados sobre series temporales de datos: (i) los basados en *funciones de proximidad*; (ii) los basados en las *características de los elementos* a clasificar y (iii) los basados en *modelos*:

- *Métodos basados en funciones de proximidad.* Utilizan un método de búsqueda de la estructura del cluster típicamente aglomerativo. Inicialmente se calcula una matriz de proximidad P de tamaño $L \times L$. Cada uno de los componentes de la matriz P_{ij} representa la similitud entre el par de objetos (x_i, x_j) . Una vez calculada la matriz de similitudes, ya no será necesario utilizar el conjunto de datos original. Las medidas de similitud utilizadas para el cálculo de la matriz son muy variadas: Distancia Euclídea [TK99], correlación entre objetos, distancia de Hamming, medidas de distancia entre cadenas [OL97], métodos de alineamiento por programación dinámica [OFC99], etc.
- *Métodos basados en las características de los elementos.* Presentan los datos mediante la vectorización de sus características. Para el proceso de clasificación se procede a aplicar directamente métodos de clustering estáticos sobre los vectores de características mencionados. En la etapa de extracción de las características de los datos cabe mencionar, entre otros muchos, los trabajos descritos en [CNPA02] y [CN02], las aproximaciones de señales biomédicas obtenidas mediante la transformada wavelet en [NCE⁺00] y [CNE⁺00] cuyas propiedades y características aparecen descritas en [Bur97], y los métodos de aproximación poligonal descritos en [Ram72] y utilizados en [KJ96], además de hacer referencia al *Capítulo 4* donde se tratará a fondo este problema.
- *Métodos basados en modelos.* En este caso, los datos a clasificar se intentan caracterizar mediante toda una amplia gama de modelos (los basados en funciones, los de autómatas, etc.). También se han utilizado técnicas de modelado basadas en métodos de regresión lineal [GS99], en redes neuronales [Kou98] y en autómatas de estados finitos (Finite State Automata, FSA) (este último método ha sido utilizado sobre todo para tareas de reconocimiento sintáctico de patrones y de reconocimiento del habla). Aquí también podríamos incluir el modelado de señales biomédicas mediante HMM.

3.6. Conclusiones

En el presente capítulo se han recopilado muchos de los algoritmos existentes en la literatura y que están implicados en las diferentes etapas de procesamiento de una señal electrocardiográfica.

Para la reducción del ruido (*Sección 3.1.1*) se han presentado algoritmos basados en técnicas clásicas de filtrado, filtros de media móvil, promedio de latidos, aproximación mediante funciones y basados en la transformada Wavelet. En general, no ofrecen buenos resultados en la aplicación a la señal electrocardiográfica, ya que además del propio ruido, eliminan puntos significativos de la señal.

Los métodos utilizados en la reducción de la interferencia de la red (*Sección 3.1.2*) ofrecen unos resultados bastante satisfactorios, ya que se trata de un problema presente en multitud de campos y cuya solución no es excesivamente compleja, principalmente debido a que la frecuencia de la interferencia es conocida.

Para la reducción de las variaciones de la línea de base (*Sección 3.1.3*), hemos visto métodos basados en filtros clásicos, adaptativos, aproximación mediante funciones, y filtros variantes en el tiempo. En general, presentan las mismas ventajas e inconvenientes que en el caso de la reducción del ruido, siendo la aproximación mediante funciones, concretamente la aproximación basada en la transformada Wavelet, la que en principio combina exactitud de los resultados con sencillez en la aplicación, aunque debe todavía adaptarse al caso particular de la señal electrocardiográfica.

En cuanto a las técnicas de extracción de las características de la señal ECG se han citado algunos de los algoritmos más interesantes y que mayores prestaciones proporcionan a la hora de comprimir las señales biomédicas como el de aproximación poligonal.

Respecto a la detección de puntos significativos en el proceso de segmentación del ECG, hemos presentado métodos basados principalmente en el tratamiento digital de la señal y en el uso de su primera y segunda derivadas.

En el apartado de selección y extracción de características de la señal hemos querido incluir referencias bibliográficas a dos de los métodos utilizados en este caso: el PCA y los HMM.

Finalmente, en el apartado dedicado al reconocimiento de patrones, hemos

distinguido entre los métodos de clasificación supervisados y los no supervisados.

Como conclusión cabría destacar que no se ha encontrado trabajo previo al respecto de la aplicación que se pretende desarrollar a lo largo de esta tesis: el procesamiento automático de una señal Holter ECG para la reducción de características de los latidos que la componen mediante la aplicación de distintos métodos estadísticos/estocásticos (concretamente PCA y HMM) y su posterior aplicación al proceso de clustering.

Parte II

Caracterización de la señal electrocardiográfica

Capítulo 4

Extracción de características

Resumen. El desarrollo y perfeccionamiento de los dispositivos de registro de señales biomédicas conlleva un incremento en la calidad de las mismas. Por ello se plantea el problema del tratamiento y almacenamiento de una elevada cantidad de información derivada de este tipo de señales. En nuestro caso y para agilizar el procesamiento del Holter se propone anticipar la fase de extracción de características mediante la aproximación poligonal de los datos originales, reduciendo de esta manera la cantidad de información que se manejará a posteriori. En este capítulo se analizan los resultados de la extracción de características mediante aproximación poligonal de la señal Holter. El estimador de calidad que nos permitirá analizar la calidad de las aproximaciones es el error residual. Nuestro propósito será el de establecer un método de aproximación eficiente que, con una tasa de error asumible, proporcione una señal con el mayor ratio de compresión posible. Además también se optimizará la velocidad de cálculo de la aproximación con el propósito futuro de comprimir los datos en tiempo real (durante la fase de adquisición del Holter). El capítulo termina con una comparativa entre los resultados obtenidos y con la presentación de las conclusiones.

Contenidos

4.1. Introducción	111
4.2. El problema de la aproximación	113
4.3. Cálculo del error	115
4.4. Definición del umbral	121
4.5. Proceso de aproximación poligonal	122
4.6. Experimentos y resultados	126
4.7. Conclusiones	133

4.1. Introducción

En los dispositivos comerciales destinados al registro portátil de señales *Holter* se pueden llegar a generar registros ambulatorios de duraciones de hasta 48 horas (ver *Capítulo 2*). Si tomamos como ejemplo un registro *Holter* con esta duración, cuyas muestras han sido adquiridas a una frecuencia de 360 Hz y con una precisión de 12 bits por muestra, obtendremos como resultado una señal con un tamaño aproximado de 89 MB. Este elevado volumen de información complica extremadamente las etapas de almacenaje y procesamiento de las señales por lo que resulta evidente la necesidad de eliminar toda aquella información redundante que no va a influir en el análisis posterior de la señal. Para ello existen una amplia gama de métodos que aproximan la señal original mediante líneas poligonales que comprimen la información de su morfología haciéndola más compacta [KJ96]. Así, todos los procesos posteriores que actúen sobre la señal comprimida se verán beneficiados de una reducción sustancial de la información a tratar, disminuyendo su duración total. Algunas de las aproximaciones que se citan como ejemplo se utilizan para la detección del pulso carotídeo aproximando la señal original mediante polinomios de grado N ($N > 1$) [PJ92], o en el campo de la electroencefalografía (EEG) donde se usa la función de autocorrelación de la señal [MH79]. La aproximación mediante líneas poligonales que contengan puntos de la señal original resulta computacionalmente sencilla de calcular y es la que vamos a utilizar en el caso de las señales electrocardiográficas.

Una señal digital estará formada por n puntos $\{(x_1, y_1), \dots, (x_n, y_n)\}$ donde x_i representa el instante de tiempo i -ésimo en el que se ha registrado el valor de señal y_i (con $1 \leq i \leq n$). El objetivo a perseguir en la aproximación se puede definir de dos formas:

1. Dado un número de puntos determinado para la aproximación, maximizar su ratio de compresión, minimizando las pérdidas de información y su error residual [PH74].
2. Dado el error máximo permitido, minimizar el número de puntos de la línea poligonal y las pérdidas de información.

Como resulta difícil el identificar la información crítica diferenciándola de aquella que no lo es, es posible que en el proceso de aproximación se pierdan algunos de los patrones de la señal, en especial si la relación señal a ruido (SNR) es pequeña. En este aspecto, la calidad de las señales electrocardiográficas suele ser elevada presentando niveles de señal (en los patrones de las ondas *QRS*, *T* y *P*) mucho mayores que el ruido, por lo que, utilizando el método de aproximación poligonal propuesto, se consiguen alcanzar con éxito todos los objetivos citados anteriormente.

A continuación se definen una serie de términos que se van a utilizar en este capítulo:

- *Información crítica*. Conjunto de puntos característicos de la señal ECG original que puedan resultar de interés para su posterior análisis e interpretación. Presentan una importante relevancia clínica. El error crítico es el que se produce con la pérdida de información crítica.
- *Información redundante*. Conjunto de muestras de la señal original que pueden aproximarse mediante una única línea poligonal sin que el error residual supere cierto umbral. El error asumible es el que se produce cuando la información perdida es redundante.
- *ECG multiregistro*. Normalmente un dispositivo Holter registra la actividad eléctrica cardíaca obtenida desde varias derivaciones generando un ECG por cada una de ellas.
- *Método de aproximación eficiente*. Aquel que no necesita de la definición de un elevado número de parámetros para su correcto funcionamiento. En el caso que nos ocupa sólo utilizaremos como parámetro el error residual.
- *Error residual*. Distancia existente entre un punto de la señal original y su aproximación poligonal. Existen distintos criterios para evaluar esta diferencia (error absoluto, error cuadrático).
- *Error residual asumible*. En la señal aproximada no existe pérdida de información apreciable al ojo del inspector humano. Se mantiene toda la información crítica. El error residual proporcionado en este caso por la aproximación sería un error asumible.

- *Error total*. Se calcula sumando los errores residuales para cada uno de los puntos de la señal original.
- *Error absoluto medio (Mean Absolute Error, MAE)*. El el valor medio del conjunto de errores absolutos calculados para cada uno de los puntos de la señal original respecto de su aproximación. El MAE se puede utilizar como estimador de calidad para la comparación entre métodos de aproximación diferentes.
- *Error cuadrático medio (Mean Squared Error, MSE)*. El el valor medio del conjunto de errores cuadráticos calculados para cada uno de los puntos de la señal original respecto de su aproximación. El MSE también se puede utilizar como estimador de calidad para la comparación entre métodos de aproximación diferentes.
- *Ratio de compresión (Compression Ratio, CRP)*. Este ratio se calcula a partir del tamaño C de la aproximación referido al de la señal original O (Ecuación 4.1) [KJ96]. Cuando hablamos de señales discretas este tamaño se suele calcular en bits (multiplicando la resolución de la muestra por el número de muestras que componen la señal).

$$CRP = \frac{O - C}{O} \quad (4.1)$$

4.2. El problema de la aproximación

La característica de la larga duración de las señales ECG las convierten en candidatas ideales para la compresión de su información mediante una aproximación de tipo poligonal.

En este caso, y teniendo en cuenta que los registros ECG constituyen un tipo de señales biomédicas en las que se representan amplitudes eléctricas en función del tiempo, está claro que los puntos quedan enmarcados dentro de un espacio bidimensional donde la abscisa representa el instante de tiempo en el que se ha obtenido la muestra eléctrica, cuyo valor de amplitud queda representado en la ordenada.

De una forma general, podemos definir el problema de la aproximación poligonal de una señal de la siguiente manera: dada la señal electrocardiográfica

con n puntos $\{(x_1, y_1), \dots, (x_n, y_n)\}$, queremos reducirla a aquellos m puntos cuya unión mediante una línea poligonal mejor la aproximen, para unos parámetros definidos (error, número de segmentos, fiabilidad, etc.). Para ello será necesario contar con una métrica que nos proporcione un indicador de calidad de cada uno de los puntos originales (x_i, y_i) respecto de la serie de segmentos con que son aproximados. Llamaremos $d\{(x_i, y_i), (x_k, y_k), (x_{i+1}, y_{i+1})\}$ a la función utilizada para medir la distancia entre el punto de la señal original (x_k, y_k) y el segmento de la señal aproximada definido por los puntos (x_i, y_i) y (x_{i+1}, y_{i+1}) . Si lo que queremos es minimizar el error total entre la señal y su aproximación de m puntos, deberemos encontrar los índices i_1, \dots, i_m de aquellos m puntos que minimicen la suma de los errores residuales (4.2).

$$\min_{\{i_1, \dots, i_m\}} \sum_{j=1}^m \sum_{x_{i_j} < x_k < x_{i_{j+1}}} d\{(x_{i_j}, y_{i_j}), (x_k, y_k), (x_{i_{j+1}}, y_{i_{j+1}})\} \quad (4.2)$$

Dependiendo de la función de distancia, el problema de la minimización puede resultar computacionalmente muy costoso y, en la mayoría de los casos, la aproximación no se puede resolver en tiempo real, lo que supone un importante handicap para los sistemas de procesamiento de señales biomédicas. Para reducir este coste y cuando se determinen las líneas poligonales de la aproximación, en vez de calcular nuevos puntos para los segmentos, se utilizarán puntos pertenecientes a la señal original, aunque de esta manera no se minimice el error residual (*Figura 4.1*). A pesar de conseguir un error total asumible, es posible que en alguno de los puntos el error residual sea lo suficientemente grande como para perder información crítica (pico de onda), aunque este error quede compensado en el total por errores despreciables en el resto de los puntos. Para evitar estas pérdidas se establece un umbral máximo para el error residual. Estas consideraciones cambian el planteamiento inicial por lo que ahora el proceso de aproximación consistirá en la selección de los puntos de la señal original que definan una línea poligonal con errores residuales inferiores al umbral y error total asumible. El problema se simplifica todavía más si se define a priori el número de puntos a utilizar en la aproximación.

Por otra parte, si de alguna forma se pudiera establecer inicialmente el umbral máximo (lo que implica un conocimiento previo de las características de la

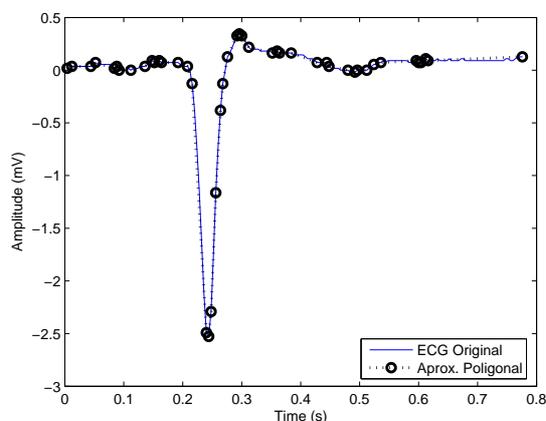


Figura 4.1: Fragmento de una señal electrocardiográfica real (línea continua) y su aproximación mediante una línea poligonal (línea punteada).

señal) sería posible el cálculo optimizado de la aproximación, independizando todo este proceso de un tipo de señales (las biomédicas) que presentan una gran variabilidad (escalado en el proceso de adquisición, aparición de la línea de base, interferencias, ruidos, etc.) (Figura 4.2). En el método de estimación del umbral utilizado en los experimentos [KJ96], se empieza examinando la naturaleza de la señal aproximando poligonalmente un fragmento de la misma y determinando la distribución estadística de los errores residuales para dicho fragmento. A partir de esta distribución se determina el umbral máximo del error residual. Actuando de esta forma se calcula un umbral particular para cada señal procesada.

4.3. Cálculo del error

Para el cálculo del error se utiliza una función de distancia. El error residual para un punto de la señal original se define como la mínima distancia existente entre ese punto y la línea poligonal que lo aproxima. Por ello el método de aproximación queda unívocamente definido por esta función. Cuando hagamos referencia a los resultados de una determinada aproxima-

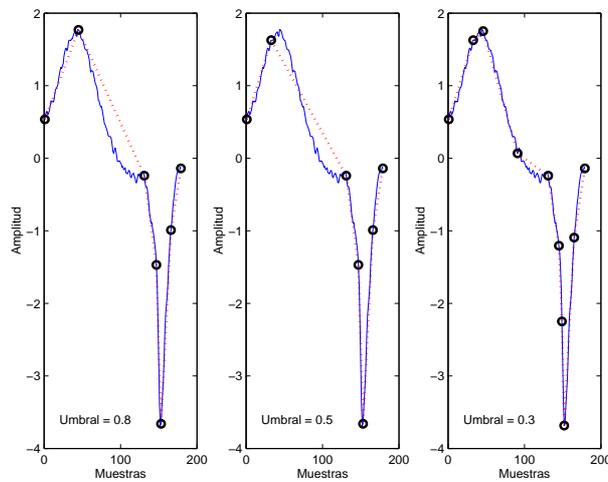


Figura 4.2: La gran variabilidad presentada por las señales biomédicas propician el cálculo particularizado de un umbral de error para cada una de ellas. El problema del ajuste del umbral se plantea en la pérdida de puntos característicos de la señal original: al disminuir el umbral (de 0,8 a 0,5) se pierde el máximo de la señal, que deja de ser un punto perteneciente a la aproximación. Por otra parte, si el umbral se hace lo suficientemente restrictivo (0,3) se vuelve a recuperar la información crítica perdida. La función de error utilizada en este caso para la aproximación ha sido el error absoluto.

ción, se hablará indistintamente del error o de la métrica utilizada en ese caso. El conjunto de funciones de distancia o errores (justificando en cada caso su idoneidad para la aproximación de las señales electrocardiográficas) que, a continuación se describen son los siguientes: *error absoluto*, *error cuadrático*, *error de superficie* y *error de amplitud*, así como la *distancia Euclídea modificada*.

Error Absoluto

Viene determinado por el valor absoluto de la diferencia entre las ordenadas del punto de la señal original seleccionado y la de su correspondiente proyección sobre el segmento utilizado en la aproximación de ese tramo de señal (Figura 4.3). La simplicidad de cálculo del error definido hace que el tiempo de procesamiento no sea excesivamente grande.

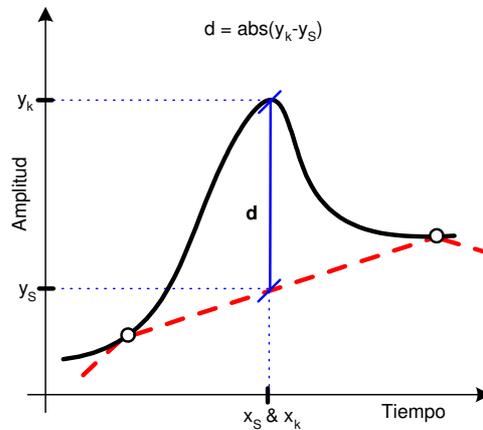


Figura 4.3: El error d se estima como el valor absoluto de la diferencia de ordenadas entre un punto de la señal original (línea continua) y su proyección (en el eje de abscisas) sobre la aproximación utilizada en ese tramo (línea discontinua). Con la referencia x_s & x_k se quiere significar la coincidencia entre las abscisas de los puntos tratados.

El problema que se plantea al utilizar esta distancia es que, en todas aquellas zonas de la señal original donde se presentan pendientes abruptas (cambios bruscos de amplitud), el cálculo de la aproximación no es capaz de eliminar la redundancia de la señal [SG80].

Existe una versión para la medida de la distancia a partir del error absoluto conocida como estimación por rango dinámico acumulado [Cue01]. En este caso se procede a la acumulación de los errores absolutos de cada punto hasta que, en cierto momento, la suma supere cierto valor umbral. Entonces se selecciona el punto en el que se ha superado el umbral para determinar un nuevo segmento de la aproximación. En la acumulación del rango dinámico la conservación de la información crítica de la señal original ni siquiera depende de la determinación del umbral, por lo que no se va a utilizar para aproximar señales electrocardiográficas.

Error Cuadrático

En este caso el error viene determinado por el cuadrado de la diferencia entre las ordenadas del punto de la señal original seleccionado y la de su correspondiente proyección sobre el segmento utilizado en la aproximación de ese tramo de señal. Tomando como referencia la *Figura 4.3* el error cuadrático se calcula como $d = (y_k - y_s)^2$. En este caso el problema que aparece es similar al comentado para el error absoluto.

Error de Superficie

En este caso se calcula el error de superficie para cada punto de la señal original a partir de la superficie del triángulo definido por dicho punto y los puntos extremos del segmento de aproximación [RR92] (*Figura 4.4*). Dos son los problemas planteados en este caso. Por un lado, la dificultad en el cálculo de la superficie de error. Por otro lado se tenemos el problema de la distorsión aparecida en los picos de señal (con la consiguiente pérdida de información crítica) y que se da en aquellas zonas en las que la señal presenta pendientes abruptas (cambios bruscos de amplitud).

Error de Amplitud

Propuesto en [KJ96], se define la distancia desde un punto al segmento de aproximación $d\{(x_i, y_i), (x_k, y_k), (x_{i+1}, y_{i+1})\}$ (*Ecuación 4.3*) como la diferencia normalizada de los productos de las diferencias de abscisas y ordenadas del punto a medir y los extremos del segmento. Al no tener que realizar ninguna comprobación previa sobre el punto considerado, el cálculo de la distancia es computacionalmente rápido. Como métricas anteriores, en las zonas de señal donde se presenten pendientes abruptas, el cálculo de la aproximación no elimina completamente la redundancia de la señal [SG80].

$$d = \frac{|(x_{i+1} - x_i)(y_k - y_i) - (y_{i+1} - y_i)(x_k - x_i)|}{x_{i+1} - x_i} \quad (4.3)$$

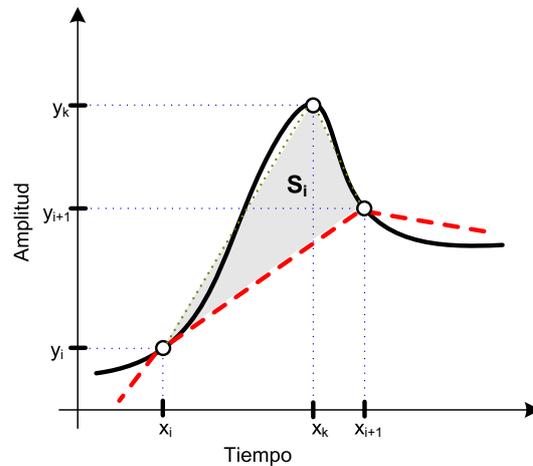


Figura 4.4: En este caso el error se estima mediante el cálculo de la superficie (sombreada) del triángulo definido si tomamos como vértice por el punto de la señal original (x_k, y_k) y como extremos los del segmento de aproximación $(x_i, y_i)(x_{i+1}, y_{i+1})$.

Error basado en la distancia Euclídea modificada

Dados los dos puntos que definen un segmento de la aproximación $((x_i, y_i)$ y (x_{i+1}, y_{i+1}) en la Figura 4.5) y una vez determinado el punto de la señal original sobre el que se va a calcular la medida del error, el error definido equivale a (i) la distancia Euclídea (los dos primeros términos en la Ecuación 4.4) entre el punto y el extremo más próximo del segmento si el punto no se encuentra localizado entre las normales a los extremos del segmento (un ejemplo de este primer caso lo constituye la distancia d_{k+1} calculada sobre el punto (x_{k+1}, y_{k+1}) en la Figura 4.5) y, en caso contrario, el error definido equivale a (ii) una modificación de la distancia Euclídea (último término de la Ecuación 4.4) entre el punto el segmento definidos por sus extremos (un ejemplo de este segundo caso lo constituye la distancia d_k calculada sobre el punto (x_k, y_k) en la Figura 4.5) [KJ96].

Como se observa en la Figura 4.5 y en la Ecuación 4.4 la distancia propues-

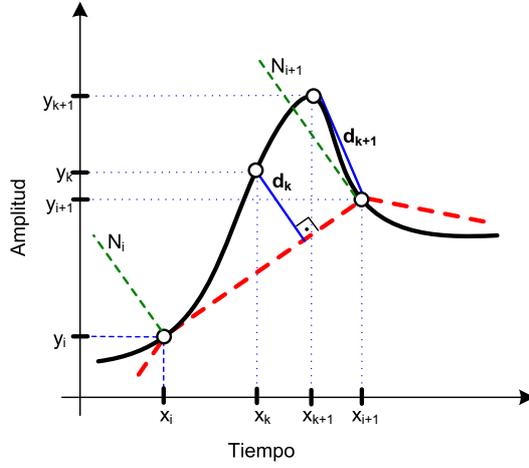


Figura 4.5: Se observan los dos casos: (i) la distancia Euclídea d_{k+1} calculada para el punto (x_{k+1}, y_{k+1}) que no se encuentra localizado entre las normales N_i y N_{i+1} de los extremos del segmento de la aproximación, y (ii) la distancia Euclídea modificada d_k calculada para el punto (x_k, y_k) que se encuentra entre las normales N_i y N_{i+1} .

ta para los puntos localizados entre las normales de los extremos de la línea poligonal $d(x_i, y_i, x_k, y_k, x_{i+1}, y_{i+1})$, se calcula a partir de la perpendicular al segmento de aproximación, d_k . En caso contrario se utiliza la distancia Euclídea al punto más próximo, d_{k+1} .

$$d = \begin{cases} \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}, & \text{si } (x_i - x_k)^2 + (y_i - y_k)^2 + \\ & + (x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 < (x_k - x_{i+1})^2 + (y_k - y_{i+1})^2 \\ \sqrt{(x_k - x_{i+1})^2 + (y_k - y_{i+1})^2}, & \text{si } (x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + \\ & + (x_k - x_{i+1})^2 + (y_k - y_{i+1})^2 < (x_i - x_k)^2 + (y_i - y_k)^2 \\ \frac{|(x_{i+1} - x_i) \cdot (y_i - y_k) + (y_{i+1} - y_i) \cdot (x_k - x_i)|}{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}, & \text{en caso contrario} \end{cases} \quad (4.4)$$

De esta forma se soluciona el problema de la aproximación en zonas abrupt-

tas de la señal (picos de señal que aparecen en la forma de onda de una señal electrocardiográfica) ya que en el pico de la onda no se toma la distancia más corta al segmento sino la Euclídea al extremo (más larga) con lo que, probablemente, se supere el umbral y se tome como nuevo punto de la aproximación el máximo de la onda. Como se deduce de la *Ecuación 4.4*, y al necesitar de una comprobación previa para conocer la situación del punto evaluado respecto de los extremos del segmento, la cantidad de operaciones matemáticas ejecutadas durante el cálculo de aproximaciones poligonales de este tipo es grande, y el tiempo de procesamiento se ve penalizado por la complejidad matemática de esta métrica.

4.4. Definición del umbral

Antes de iniciar el algoritmo que se describirá en 4.6 es necesario calcular el umbral ε que se utiliza para determinar si el error residual es asumible o por el contrario si va a suponer que el punto medido determine un nuevo segmento en la aproximación. Este umbral no se puede calcular a priori debido a las distorsiones que la señal electrocardiográfica sufre durante el proceso de adquisición (nivel de continua, alteraciones de la línea de base, ruido, etc.) y que hacen que varíen sus características. Para obtener una estimación inicial ε , necesitaremos conocer la función de distribución de los errores residuales. Para ello se propone inicialmente una aproximación poligonal de la señal determinando aleatoriamente la longitud de los segmentos de la aproximación, dentro de unos límites de longitud. A continuación se seleccionan también aleatoriamente m puntos del tramo de señal original aproximada para, a partir del error elegido, generar un vector de errores residuales R (*Ecuación 4.5*). Este proceso se repite hasta considerar que, con el número de experimentos realizados, se obtiene una distribución del error residual lo suficientemente representativa para la señal a aproximar (*Figura 4.6*).

$$R_i = d\{(x_i, y_i), (x_{i+r_1}, y_{i+r_1}), (x_{i+r_1+r_2}, y_{i+r_1+r_2})\}, \quad i = 1, \dots, m. \quad (4.5)$$

Donde $2 \leq r_1 \leq l_1$, $2 \leq r_2 \leq l_2$ y r_1, r_2 son números aleatorios uniformemente distribuidos y l_1, l_2 determinan los límites superiores de dichos

números.

Asumiendo que la distribución del error obtenida aleatoriamente será similar a la del proceso de aproximación real y, a la vista de la misma, se ordenan crecientemente los errores obtenidos determinando un valor porcentual P que nos permitirá seleccionar como valor umbral ε el error residual (del total de errores del vector R_i) que corresponda con la probabilidad definida $R(\frac{P \cdot m}{100})$ (Figura 4.6, donde se especifica el valor del umbral obtenido para las diferentes distribuciones de error según la métrica utilizada). Además, este porcentaje supone una estimación inicial del índice de compresión que se va a obtener en la aproximación ya que, durante el proceso, una media de P puntos originales serán aproximados por un mismo segmento al presentar un error residual inferior al umbral. También existen otros métodos para la estimación del umbral que han sido desechados debido al elevado coste computacional presentado. La estimación de los valores (l_1, l_2) que limitan la longitud de los segmentos aleatorios (r_1, r_2) , dependerá de la naturaleza de la señal y de su frecuencia de muestreo. Una aproximación válida para señales electrocardiográficas utilizada en la práctica es la de $l_1 = l_2 = 50/(100 - P)$ [KJ96], de tal forma que mediante $l_1 + l_2$ se estimen los valores medios del factor de compresión y de la longitud de los segmentos utilizados en la aproximación.

4.5. Proceso de aproximación poligonal

Dada la señal original definida por los puntos $\{(x_1, y_1), \dots, (x_n, y_n)\}$, el proceso de aproximación se inicia determinando el primero de los segmentos. Para ello se utiliza un subconjunto de puntos consecutivos, de tamaño k_1 , tomados del grupo de puntos que conforman la señal original. A partir de este subconjunto se traza un primer segmento de aproximación intentando minimizar el error residual medido. El cálculo de la recta que minimice el error puede resultar computacionalmente complejo por lo que, en nuestro caso, la línea poligonal de aproximación estará determinada por algunos de los puntos pertenecientes a la señal original. Este primer segmento se define mediante la línea que pasa por el primer y último punto del subconjunto de k_1 puntos seleccionados. A continuación, se calcula el desde todos los puntos a este primer segmento. Si ninguno de los errores residuales obtenidos supera valor del um-

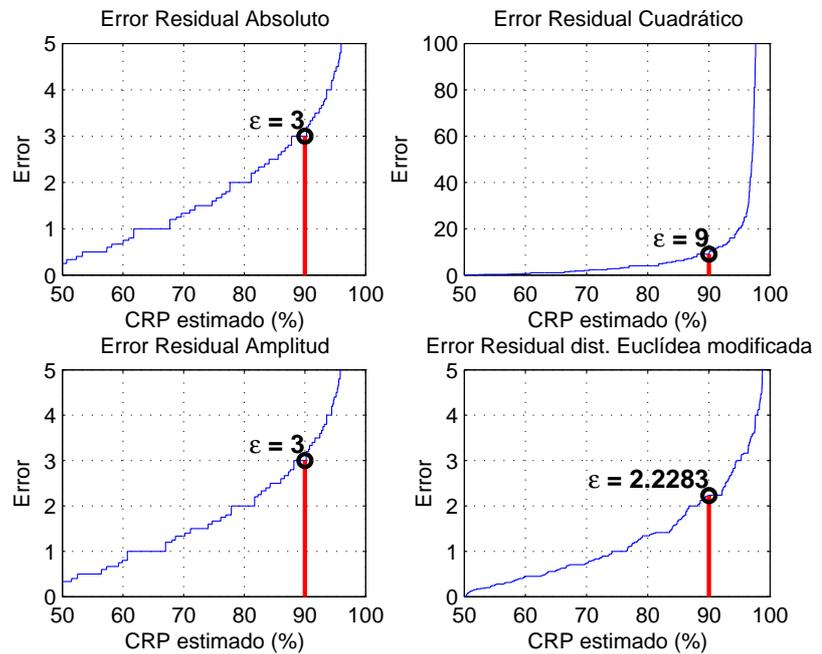


Figura 4.6: Distribución de los errores residuales obtenidos al aproximar aleatoriamente un ECG real. Sobre una misma señal se han realizado 5000 aproximaciones para cada una de las cuatro métricas utilizadas, determinando una longitud máxima del segmento de 5 muestras y un ratio de compresión aproximado del 90%. El valor del umbral ε obtenido en cada caso es el que se muestra en la figura. Como se puede apreciar, todas las distribuciones del error obtenidas con las diferentes métricas presentan una morfología similar donde los mayores errores responden al caso en el que el pico de la onda R ha sido aproximado por un único segmento. Errores medios corresponden a una mala aproximación de las demás ondas, mientras que el resto de errores son muy pequeños y producen el efecto de filtrado paso bajo del ECG (Figura 4.7).

bral definido ε , se añaden nuevos puntos al subconjunto repitiendo el proceso hasta que el error residual de alguno de los puntos supere el umbral de error ε . Éste punto se utiliza para definir un nuevo segmento de la aproximación, recalculando las distancias de los puntos del conjunto a la nueva línea poligonal y comprobando si alguno de los errores supera ε . Si se supera el umbral

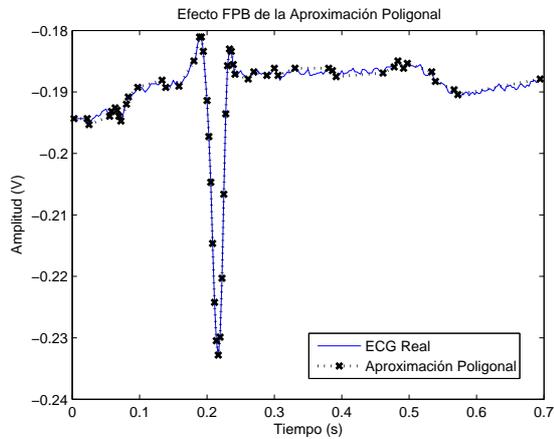


Figura 4.7: La aproximación mediante líneas poligonales (utilizando como estimador el error de amplitud) de una señal ECG real produce un efecto de filtrado paso bajo que elimina los pequeños picos de ruido.

en alguno de los puntos se repetiría el proceso de selección del punto y nuevo cálculo de la línea poligonal y de las distancias hasta conseguir que ninguno de los errores residuales supere el umbral.

Para iniciar la siguiente iteración tomaremos la abscisa del último segmento procesado j como abscisa de inicio del nuevo subconjunto de puntos a aproximar, cuyo nuevo tamaño k_i (donde i representa el número de la iteración) será igual a la longitud del último de los segmentos procesados en la anterior iteración. De esta forma, se repetiría el proceso hasta llegar al final de la señal. Los puntos (x, y) de la aproximación se determinan con las abscisas obtenidas en cada iteración y cuya proyección sobre la señal original proporciona los correspondientes valores de ordenadas. El algoritmo descrito es el que se presenta a continuación:

- 1) Inicializar $i = 1, j = 1, k = 2$
- 2) Actualizar $j = j + k$
- 3) Si $j \geq n$ finalizar
- 4) Sea p el punto que maximiza
 $D = d\{(x_i, y_i), (x_p, y_p), (x_j, y_j)\}, \quad i < p < j$
- 5) Si $D \leq \varepsilon$, ir a paso 2
- 6) $j = p$
- 7) Sea p el punto que maximiza
 $D = d\{(x_i, y_i), (x_p, y_p), (x_j, y_j)\}, \quad i < p < j$
- 8) Si $D \leq \varepsilon$, ir a paso 6
- 9) $k = j - i$
- 10) Actualizar $i = j$, ir a paso 2

(4.6)

Con este algoritmo obtenemos una línea poligonal que, dependiendo de la métrica utilizada en los pasos 4 y 7, y del umbral de error estimado en la Sección 4.4, se ajustará de una u otra forma a la señal original. Para optimizar la velocidad de procesamiento del algoritmo, resulta importante el realizar una estimación precisa de la longitud del próximo segmento a calcular ya que, mediante una buena estimación de su longitud (Ecuación 4.7), se disminuyen sensiblemente el número de iteraciones consistentes en aumentar, en cada una de ellas, el número de puntos de la señal original considerados, al no superar ninguno de ellos el umbral de error.

$$k(i+1) = \frac{i}{i+1} \cdot k(i) + \frac{1}{i+1} \cdot L \quad (4.7)$$

Donde la longitud estimada para el segmento $k(i+1)$ en la iteración $i+1$ depende de la longitud estimada $k(i)$ para el segmento procesado en la iteración i , y de la longitud real L de este mismo segmento. El número de iteraciones coincide con el número de segmentos calculados.

Además del propuesto, existen otros métodos para estimar el tamaño del conjunto de puntos a procesar (cadenas de Markov, redes neuronales, etc.) cuya complejidad de cálculo limita los beneficios en cuanto a velocidad de procesamiento se refiere.

4.6. Experimentos y resultados

El principal objetivo de los experimentos propuestos es el de determinar una métrica que proporcione la aproximación poligonal de la señal sin pérdida de información crítica y con las mejores prestaciones posibles. Los resultados de la comparativa entre la distancia euclídea modificada y el error de amplitud se pueden consultar en [KJ96]. En nuestro caso se van a presentar resultados (velocidades de procesamiento y calidad de la aproximación) utilizando cada una de las cuatro distancias citadas en la *Sección 4.3* (no se van a realizar cálculos sobre el error de superficie). En la mayoría de las pruebas se han utilizado un total de cinco señales electrocardiográficas de 30 minutos de duración seleccionadas de la base de datos del MIT [MM98] siguiendo un criterio de variabilidad. Entre estas señales queda recogida la mayor parte de la casuística existente (en cuanto a tipos de latido y ritmos cardíacos distintos se refiere) (ver www.physionet.org/physiobank/database/html/mitdbdir/tables.htm). Las señales han sido muestreadas con una frecuencia de 360 Hz, una resolución de 11 bits y un rango de tensión de ± 5 mV donde un valor de muestra de 1024 equivale a una tensión de 0 V. Para eliminar las interferencias de la red eléctrica las señales han sido previamente tratadas con un filtro digital FIR paso-bajo de 60 coeficientes con una frecuencia de corte de 50 Hz. Además, gracias a este filtrado se consigue disminuir el error total de la aproximación poligonal ya que, al suavizar la señal original, los segmentos de la aproximación se ajustan mejor a la forma de onda de la señal filtrada. Todas las pruebas se han ejecutado en una computadora equipada con un microprocesador *Intel® Pentium® IV* a 3,00 GHz y 1 GB de memoria RAM. Para los desarrollos software se ha utilizado la aplicación *Matlab®v7*.

A continuación se presentan los resultados relativos a las pruebas realizadas sobre: (i) *velocidad de procesamiento en el cálculo del umbral de error*, (ii) *velocidad de procesamiento en la aproximación poligonal*, (iii) *relajamiento del umbral de error en la aproximación poligonal*, (iv) *MSE*, (v) *MAE*, (vi) *CRP*, (vii) *Fidelidad* y (viii) *incremento de la velocidad de procesamiento en señales comprimidas*.

(i) Velocidad de procesamiento en el cálculo del umbral de error

El motivo de esta prueba es el de determinar el algoritmo más rápido en el proceso de cálculo del umbral de error (ver *Sección 4.4*). El valor de CRP utilizado en este caso ha sido del 90 % ($P = 90$) con lo que se obtienen $l_1 = l_2 = 5$. Se han procesado 5000 segmentos para obtener los valores de ϵ que aparecen en la *Figura 4.6*. Los resultados de velocidad de procesamiento son los que se detallan en la *Figura 4.8*.

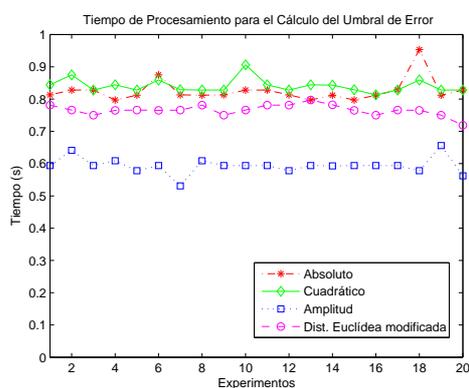


Figura 4.8: Comparativa para el tiempo de procesamiento entre las diferentes métricas descritas en la *Sección 4.3* durante el cálculo del umbral de error. Como se observa, las métricas más rápidas y las que presentan una mayor regularidad estadística son las basadas en el error de amplitud y en la distancia euclídea modificada utilizada por Koski [KJ96]. Con el fin de proporcionar cierto peso estadístico, se han efectuado un total de 20 pruebas para cada métrica.

(ii) Velocidad de procesamiento en la aproximación poligonal

En este caso se han seleccionado cinco señales electrocardiográficas procedentes de la base de datos del MIT [MM98] (señales 100, 118, 207, 212 y 217 compuestas por 650000 muestras cada una de ellas) para su aproximación poligonal mediante el algoritmo presentado en la *Sección 4.5* y utilizando las diferentes métricas presentadas en la *Sección 4.3*. Los resultados obtenidos para

la velocidad de procesamiento de las señales son los que se muestran en la *Figura 4.9*.

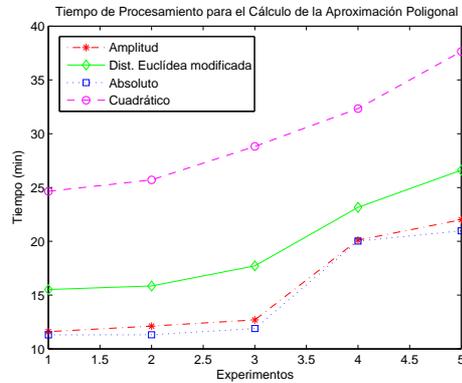


Figura 4.9: Comparativa para el tiempo de procesamiento entre las diferentes métricas descritas en la *Sección 4.3* durante las aproximaciones poligonales realizadas. Tanto la estimación del error por amplitud como por error absoluto proporcionan resultados similares que optimizan el tiempo de procesamiento respecto al resto de criterios.

(iii) Relajamiento del umbral de error en la aproximación poligonal

Tal y como se propondrá en la *Sección 5.7* dedicada a las conclusiones derivadas de las pruebas, la velocidad de procesamiento de las señales se puede mejorar relajando el umbral de error calculado. En este caso, la prueba se realiza sobre un ECG cuyo umbral de error se ve incrementado en un 500 %. La aproximación poligonal obtenida en este caso no deja de ser fiel a la forma de onda (*Figura 4.10*) incrementando la velocidad de procesamiento de la señal en un 52 % aunque a costa de aumentar el MAE en un 174 % y el MSE en un 294 %.

(iv) Error absoluto medio

Obtenidas las diferentes aproximaciones poligonales (en dos etapas, primero se define el umbral de error para un CRP del 90 % y para cada una de

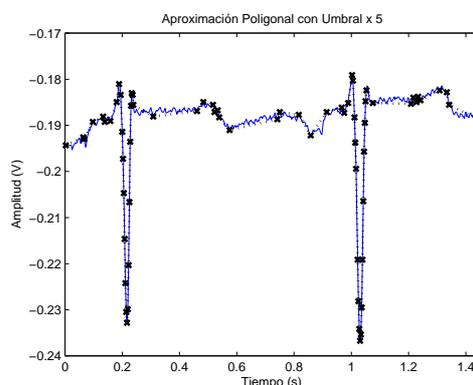


Figura 4.10: Aproximación poligonal realizada relajando el umbral de error al aumentarlo en un 500 %. Este hecho permite incrementar la velocidad de procesamiento de la señal en un 52 % aunque el MAE y MSE se ven penalizados. A pesar de todo la aproximación continúa guardando fidelidad con la señal original.

las señales analizadas y, posteriormente, se ejecuta el algoritmo de aproximación mostrado en la *Sección 4.5* sobre cada una de las métricas seleccionadas), procederemos al cálculo del error absoluto medio definido en la *Sección 4.1* para establecer una comparativa entre las diferentes métricas utilizadas. Los umbrales de error se muestran en el *Cuadro 4.1*. El MAE obtenido en cada caso aparece en la *Figura 4.11*. Como el interés de esta prueba reside en la comparación de los resultados ofrecidos por métricas diferentes, las unidades utilizadas para su cálculo han sido directamente los valores de las muestras (11 bits de resolución) sin haberlos traducido a valores de tensión.

(v) Error cuadrático medio

Al igual que en el error anterior, se calcula el error cuadrático medio definido en la *Sección 4.1*. Los umbrales de error utilizados en este caso (y que se aplican al proceso de aproximación poligonal y no al cálculo del estimador del error final) son los mismos que los utilizados en el caso anterior (*Cuadro 4.1*). Los resultados del MSE son los mostrados en la *Figura 4.12*. Como en el caso anterior y residiendo el interés de la prueba en la comparación de los resulta-

Cuadro 4.1: Umbrales de error obtenidos para la aproximación poligonal de las diferentes señales procedentes de la base de datos del MIT [MM98] y calculados con las distintas métricas. El CRP estimado es del 90% habiendo realizado un total de 1000 iteraciones por umbral calculado.

	Absoluto	Cuadrático	Amplitud	Euclídea modificada
ECG 100	3.00	3.33	3.00	3.40
ECG 118	9.60	9.00	10.0	9.00
ECG 207	3.75	3.66	3.80	3.80
ECG 212	4.40	4.60	5.00	4.80
ECG 217	4.40	4.50	4.60	4.50

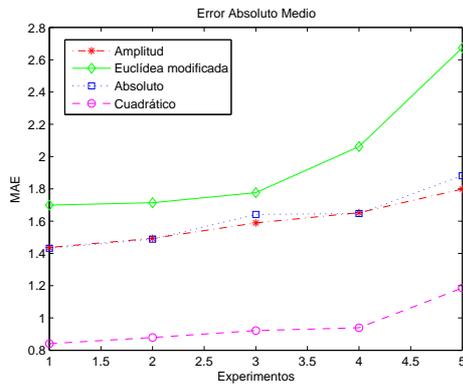


Figura 4.11: Los resultados muestran el valor del error absoluto medio de la aproximación poligonal de la señal mediante las distancias de error especificadas. Los umbrales de error utilizados son los que aparecen en el *Cuadro 4.1*.

dos ofrecidos por métricas diferentes, las unidades utilizadas para su cálculo han sido directamente los valores de las muestras sin haberlos traducido a valores de tensión.

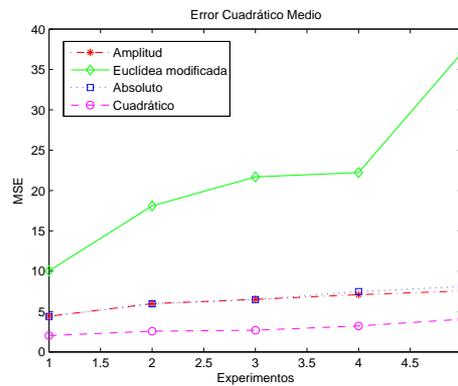


Figura 4.12: Los resultados muestran el valor del error cuadrático medio obtenido con las métricas especificadas calculado sobre las fuentes citadas procedentes de la base de datos del MIT. Los umbrales utilizados aparecen en el *Cuadro 4.1*.

(vi) Ratio de compresión

En este caso se procede a la representación del ratio de compresión obtenido para cada una de las aproximaciones poligonales (*Figura 4.13*) y cuyo cálculo queda definido en la *Sección 4.1*. Los umbrales de error utilizados son los presentados en la *Tabla 4.1*.

(vii) Fidelidad

Para la correcta evaluación de los diferentes algoritmos de aproximación poligonal resulta importante el estimar de alguna forma cuál de los algoritmos es capaz de aproximar la señal original recogiendo la información proporcionada por sus puntos característicos (en el caso de un latido normal que ha sido registrado desde las derivaciones I o II, aparecen la onda *P*, el complejo *QRS* y la onda *T*) de una forma más fidedigna. Para ello se ha procedido a etiquetar manualmente estos puntos sobre la señal original y se ha establecido un nuevo parámetro estimador que hemos denominado fidelidad. La fidelidad de una aproximación poligonal expresa porcentualmente la cantidad de puntos característicos de la señal original que han sido efectivamente con-

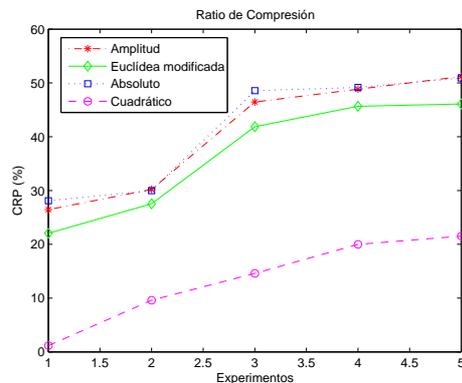


Figura 4.13: Los resultados muestran el ratio de compresión obtenido con las métricas especificadas para cada una de las fuentes referidas anteriormente.

templados en su aproximación. Los umbrales de error utilizados son los presentados en el *Cuadro 4.1*. Visualmente, los puntos característicos de la señal quedan adecuadamente aproximados en la mayoría de los casos por cada una de las métricas especificadas (ver *Figura 4.7* en la que se ha utilizado como estimador el error de amplitud con valor un umbral de 7.2, que no es demasiado restrictivo, detectando en cambio todos los puntos característicos de la señal ECG: onda *P*, complejo *QRS*, etc.). En todos los casos probados la fidelidad de la aproximación se aproxima al 100 % presentando con un porcentaje de pérdidas máximo del 1 % en la detección de los puntos característicos.

(viii) Incremento de la velocidad de procesamiento en señales comprimidas

En esta última prueba se pretende justificar la actuación preferente sobre la aproximación poligonal antes que hacerlo sobre la señal original. Esta actuación se justifica en términos de velocidad de procesamiento. Para ello se establece una comparativa entre los tiempos de procesamiento calculados para la eliminación de las variaciones en línea de base tanto sobre las señales originales como sobre sus aproximaciones poligonales. La operación seleccionada para esta prueba no resulta trivial ya que implica la utilización de la transformada *wavelet* implementando una descomposición de niveles 6 a 10 con una

wavelet madre de tipo *Daubechies 4* (Capítulo 3, Sección 3.1.3). A través de los resultados derivados de una tarea compleja de procesamiento de señal se pretende justificar la utilización de la aproximación poligonal en cualquier otro tipo de tarea a aplicar sobre el ECG. El conjunto de señales utilizadas en las pruebas han sido seleccionados de la base de datos del MIT. Los resultados obtenidos son los que se pueden observar en la *Figura 4.14*.

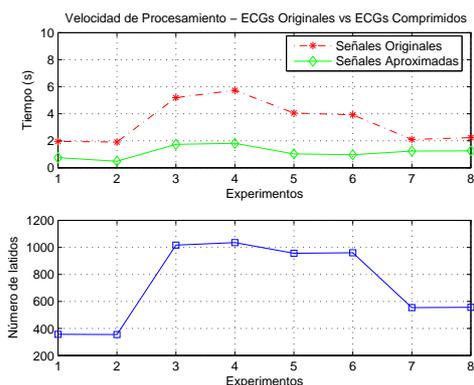


Figura 4.14: Incremento de la velocidad de procesamiento en señales comprimidas donde se puede apreciar como, sobre un conjunto de 8 señales analizadas, la mejora del rendimiento puede llegar a ser hasta del 82.5% (experimento número 7). En segundo término y para justificar el tiempo de procesamiento en cada caso aparece representado el número de latidos utilizado en los experimentos.

4.7. Conclusiones

Las pruebas anteriores han sido realizado para determinar el algoritmo óptimo de aproximación poligonal a utilizar durante el proceso de compresión de señales ECG multiregistro de tipo Holter. El motivo de esta cuidada elección es el de obtener señales ECG de reducido tamaño que contengan el máximo de información posible para la obtención de los mejores resultados en el posterior proceso de clustering de latidos.

Las conclusiones derivadas de la interpretación de los resultados obtenidos en la *Sección 4.6* son las siguientes:

1. *Velocidad de procesamiento en el cálculo del umbral de error:* donde la distancia de error que mejores prestaciones ofrece en este caso es el error de amplitud, cuya velocidad en el cálculo del umbral de error es del orden de 0.2 s inferior (un 25 % más rápido) al del resto de medidas propuestas. De todas maneras, el cálculo del umbral se ha realizado sobre un fragmento de señal compuesto por 15000 muestras, con lo que los tiempos estimados en cualquiera de los casos no resultan relevantes en el global del proceso de aproximación poligonal.
2. *Velocidad de procesamiento en la aproximación poligonal:* ya que, tal y como se presenta en la *Figura 4.9*, las distancias que mejores resultados proporcionan en este caso son las que utilizan como criterio de aproximación los errores de amplitud y absoluto. La distancia euclídea modificada presenta resultados algo peores aunque el estimador más lento es el que utiliza el error cuadrático. En la *Figura 4.15* se establecen los porcentajes de mejora del resto de medidas respecto al error cuadrático, donde se puede observar como los estimadores que utilizan los errores de amplitud y absoluto ofrecen porcentajes de mejora de entre un 40 % y un 50 % sobre el tiempo de cálculo utilizado por el error cuadrático. En el caso de la distancia euclídea modificada la mejora se establece en torno a un 25 %. Cuando hablamos del procesamiento de señales de duración media (como es el caso, en las pruebas cada señal procesada está compuesta por 650000 muestras), una mejora del 50 % se traduce en diferencias de hasta 15 minutos de retraso en el procesamiento de una misma señal dependiendo de la medida de error utilizada.
3. *MAE y MSE:* de los resultados de la *Sección 4.6* para el cálculo de los distintos errores medios (*Figuras 4.11, 4.12*) y de la *Tabla 4.1* de los umbrales de error se deriva que, una vez establecido el ratio de compresión estimado a priori y calculado el umbral de error (para cada medida del error y para cada señal y con el ratio especificado) los estimadores del error medio MAE y MSE obtenidos a posteriori sobre la aproximación son óptimos si se utiliza como métrica medida del error el error cuadrático medio, resultando bastante elevados si se utiliza como estimador el error por distancia euclídea modificada. Para los errores de amplitud y absoluto el MSE y el MAE se mantienen dentro de un rango de calidad aceptable. Esto se puede explicar si tenemos en cuenta cómo la

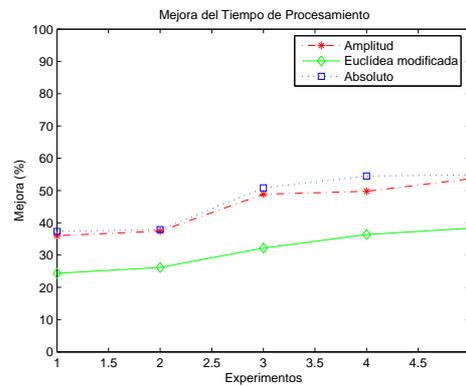


Figura 4.15: Porcentaje de mejora respecto al error cuadrático para el tiempo de procesamiento durante las aproximaciones poligonales realizadas para las señales citadas en la correspondiente prueba de la Sección 4.6.

modificación de la distancia euclídea aproxima los picos de la señal con muchos menos segmentos (y por tanto con mayor error) que el resto de métricas, aunque luego utilice un mayor número de segmentos en la aproximación de zonas ruidosas, perdiendo eficiencia en la compresión y disminuyendo el efecto de filtrado paso bajo de la señal en este tramo.

4. **Ratio de compresión:** los ratios obtenidos en la Sección 4.6 se encuentran entre el 20 % y el 50 % en el caso de utilizar como estimadores los errores de amplitud o absoluto o la distancia euclídea modificada y quedando descartado el error cuadrático cuyo CRP en cualquiera de los casos es un 20 % inferior al proporcionado por el resto. Este hecho resulta coherente con las conclusiones derivadas del apartado dedicado al cálculo del tiempo de procesamiento (donde el error cuadrático ofrece los peores resultados, ver Figuras 4.15 y 4.9) y del apartado anterior donde se realiza la comparativa entre los errores medios obtenidos (en el que es precisamente el error cuadrático el estimador que menor error medio proporciona) y todo ello debido a la utilización de un mayor número de segmentos en la aproximación poligonal.

5. **Fidelidad:** tras inspeccionar visualmente las señales electrocardiográficas reales y compararlas con sus aproximaciones poligonales se llega a la conclusión de que, para los patrones de forma de onda presentados, todas las métricamedidas de error ofrecen un coeficiente de fidelidad cercano al 100 % (para los umbrales calculados ninguna de las distancias probadas supone pérdida de información crítica). Si comparamos visualmente los resultados obtenidos entre las aproximaciones poligonales por las diferentes métricamedidas del error utilizadas, y para todas las señales aproximadas, es la distancia euclídea modificada la que mejor aproxima tanto las zonas donde la señal presenta variaciones bruscas de amplitud (por ejemplo, pico en el complejo *QRS* aproximado con un menor número de segmentos, donde el resto de distancias proporcionan más información redundante) como las zonas en las que la señal varía más suavemente (ondas *T* y *P*, donde la distancia euclídea utiliza más líneas poligonales frente a una aproximación algo más tosca realizada por el resto de métricas). Por contra tenemos que esta métrica, proporciona un CRP real del orden de un 5 % inferior al ofrecido por otras medidas (*Figura 4.13*).
6. **Incremento de la velocidad de procesamiento en señales comprimidas:** de los resultados obtenidos en el experimento de la *Sección 4.6* (*Figura 4.14*) se desprende que resulta más rápido el procesamiento de la aproximación poligonal de una señal, optimizando de esta forma la velocidad de cálculo.

Comparando los resultados de las diferentes pruebas realizadas resulta patente la superioridad de prestaciones que, sobre el resto de medidas del error, presentan estimadores simples como el del error de amplitud o el error absoluto. Si a todo esto añadimos que la fidelidad de todas las aproximaciones es elevada (manteniendo hasta el 99 % de la información crítica del ECG), se puede concluir con que es el estimador por error absoluto (que mejora levemente las prestaciones ofrecidas por el error de amplitud, en cuanto a tiempo de procesamiento y CRP se refiere) el que optimiza el proceso de cálculo de la aproximación poligonal de señales ECG de larga duración, permitiendo las aproximaciones sin pérdida de información crítica y obteniendo ratios de compresión del 50 % si el umbral de error es mínimo y de hasta el 76 % con un

umbral de error relajado en un 500 %.

La justificación de una buena aproximación poligonal se demuestra en el último de los experimentos realizados en la *Sección 4.6* donde se comprueba como la velocidad de procesamiento de las señales comprimidas resulta incrementada.

Capítulo 5

Preprocesamiento y segmentación

***Resumen.** Una vez adquirida la señal ECG mediante el dispositivo Holter, deberemos prepararla adaptando sus características para su posterior tratamiento informático. En nuestro caso se ha procedido a la aplicación de diversas técnicas tanto para la eliminación de la interferencia de la red como para la corrección de la línea de base. La eliminación del ruido se consigue a través de la extracción de las características de la señal mediante aproximación poligonal. Además, también se ha incluido el proceso de segmentación de los latidos. En la vertiente práctica se presentan los resultados obtenidos de la comparativa realizada para todas las operaciones de preprocesado entre las señales ECG Holter originales y sus correspondientes aproximaciones. Finalmente se proponen las conclusiones derivadas de los experimentos desarrollados.*

Contenidos

5.1. Introducción	141
5.2. Interferencia de la red eléctrica	141
5.3. Variaciones de la línea de base	142
5.4. Ruido interferente	143
5.5. Segmentación de la señal en latidos	144
5.6. Normalización de los latidos	147
5.7. Conclusiones	148

5.1. Introducción

Tal y como ya hemos comentado, tras la adquisición de la señal ECG, tenemos que prepararla adaptando sus características para un posterior tratamiento informático (en nuestro caso el clustering de los latidos que la forman). Esto es debido a que la señal obtenida presenta una serie de elementos que le son ajenos (ruido, interferencia de la red de alimentación o variaciones de la línea base) y que enmascaran la información, por lo que deberán ser aislados y minimizados para que, en las siguientes fases de análisis, los resultados sean independientes de todos estos factores. La eliminación de la interferencia de la red y la corrección de la línea de base se llevarán a cabo inmediatamente después de la adquisición de la señal desde el Holter. Para la eliminación del ruido y para la segmentación se ha optado por la previa compresión de la señal ECG (*Capítulo 4*). Además, y con el objetivo final del clustering de los latidos de la señal, también incluiremos en este capítulo el proceso de segmentación desarrollado para la identificación de los latidos del Holter ECG. Las señales utilizadas, sus características y su composición son las descritas en el *Capítulo 2*.

5.2. Interferencia de la red eléctrica

En todos los dispositivos electrónicos y debido a la red de alimentación eléctrica, aparece una señal interferente localizada en una frecuencia determinada. En la red eléctrica europea, la interferencia se localiza en los 50 Hz, mientras que en los Estados Unidos la interferencia está sobre los 60 Hz. En nuestro caso, si el dispositivo de adquisición de datos no es capaz de eliminar la interferencia introducida por la red de alimentación durante el muestreo de la señal, será necesario eliminarla a posteriori mediante la operación de filtrado correspondiente. Para ello, todas las señales utilizadas en este trabajo han sido procesadas por un filtro elimina-banda centrado bien en la frecuencia de 50 Hz (para los ECG adquiridos en europa), bien en la de 60 Hz (para los ECG del MIT).

5.3. Variaciones de la línea de base

Las variaciones de la línea base suponen una interferencia de baja frecuencia y de cierta amplitud que debe ser reducida (e idealmente eliminada) para no alterar el resultado del procesado posterior (*Figura 3.7*). En nuestro caso utilizaremos un método basado en la transformada Wavelet [CNE⁺00]. Si consideramos la capacidad de las wavelet para suprimir un polinomio, dependiendo de los momentos nulos que éste posea, podremos aprovechar esta propiedad para reducir las variaciones de la línea base de la señal ECG. En concreto, se utiliza una descomposición de niveles 6 a 10 con una wavelet madre de tipo *Daubechies 4*. Los resultados obtenidos de la eliminación de la línea base sobre un latido son los que se presentan en la *Figura 5.3*.

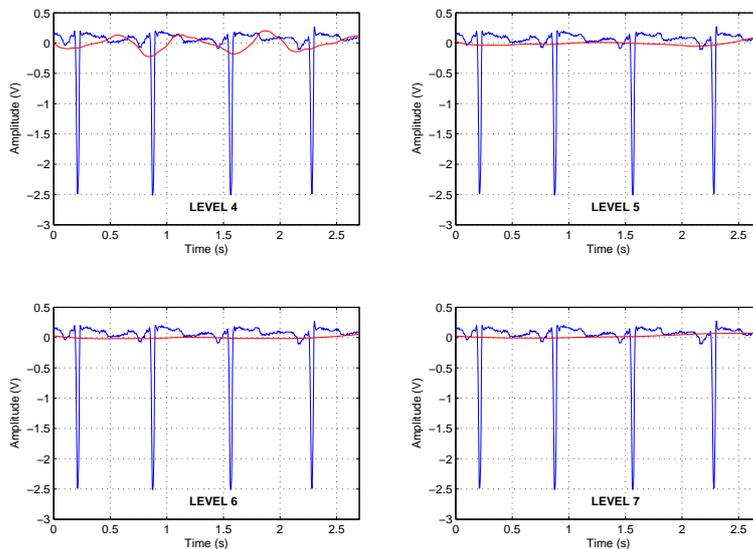


Figura 5.1: Aproximaciones sucesivas para la corrección de la línea base en una secuencia de latidos.

Evidentemente, el proceso de eliminación debería mejorar (en términos de consumo de CPU) al utilizar como fuente la señal comprimida (menor número

de muestras) en lugar de la original. Tal y como se puede observar en la *Figura 5.4* la comparativa obtenida para la operación de eliminación de la línea de base entre las señales originales y sus aproximaciones poligonales no mejora las prestaciones tanto como cabría esperar. De un total de 49 señales con 49318 latidos y 10790685 muestras procesadas para las fuentes originales sobre las 3128848 muestras procesadas en sus respectivas aproximaciones (un 70% menos de información a procesar), únicamente se obtiene una mejora del rendimiento del procesador del 5%. Este resultado puede ser debido en parte a la penalización temporal que supone el hecho de tener que guardar el vector con la información temporal en el caso del fichero de la aproximación poligonal.

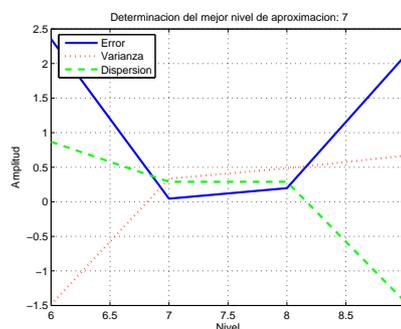


Figura 5.2: Cálculo del mejor nivel para la aproximación de la línea de base mediante wavelets. En este caso el valor mínimo del error proporciona la mejor aproximación en el nivel 7.

5.4. Ruido interferente

Para realizar distintos los experimentos objeto del presente trabajo se han elegido señales ECG con una buena relación señal a ruido (SNR) por lo que no ha sido necesario ningún tipo de filtrado al respecto. Además, al reducir el volumen de información aproximando poligonalmente la señal original se consigue el efecto de su filtrado paso bajo, debido principalmente a que el proceso de compresión supone la minimización de la información redundante (entre la

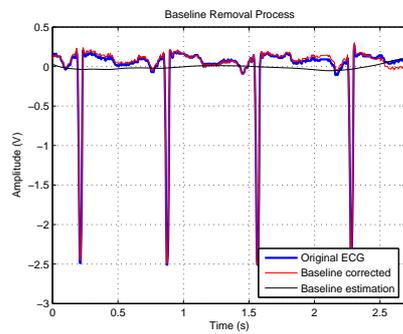


Figura 5.3: Estimación de la línea de base y posterior corrección de la señal ECG.

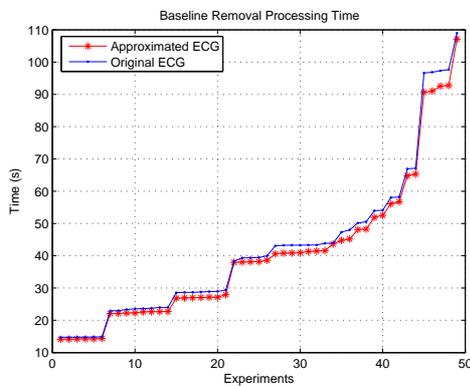


Figura 5.4: Comparativa entre las prestaciones obtenidas durante el proceso de eliminación de la línea de base entre la señal Holter original y su aproximación poligonal.

que se encuentra el ruido) manteniendo únicamente la información de interés para el análisis posterior de la señal (puntos característicos) [Kos96].

5.5. Segmentación de la señal en latidos

En esta etapa se implementa el segmentado de la señal por latidos, es decir, se extraen y contabilizan los latidos del ECG. Mediante un bucle, se recorre el

registro ECG detectando los puntos característicos de cada uno de los latidos (que dependerán de su morfología) para determinar de esta manera el inicio y fin de cada uno de los latidos que lo componen y almacenarlos finalmente en memoria, de forma independiente. El problema de la segmentación del ECG implica el conocimiento previo de la forma de onda tratada para poder detectar con precisión sus puntos característicos [Hor75]. Los puntos más significativos del registro ECG son los correspondientes a los picos de las ondas Q , R , S , así como los que determinan las características de los complejos P y T y los segmentos PQ , ST y TP . En muchos de los casos, y para simplificar el proceso, sólo es necesario identificar el pico de la onda R (Figura 5.5), calculando la frecuencia cardíaca como la inversa del tiempo que transcurre entre la aparición de dos ondas de tipo R . El número de latidos también se obtiene a partir del número de picos detectados.

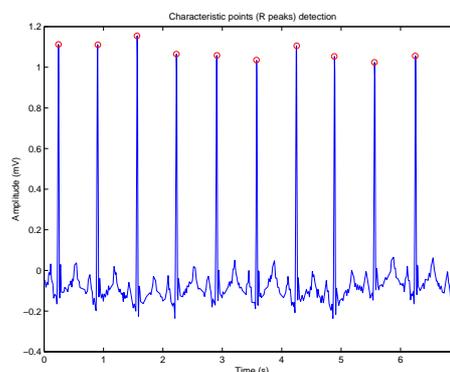


Figura 5.5: Detección de los puntos característicos (pico de la onda R) de los latidos (etiquetados como normales) del ECG 205 de la base de datos *mitdb*.

Los ECG utilizados provienen de la base de datos del MIT, cuyos latidos han sido previamente etiquetados mediante una inspección visual realizada por profesionales. La herramienta desarrollada en nuestro caso para proceder al etiquetado de los latidos segmentados es la que aparece en la Figura 5.6. Si representamos las muestras resultantes del proceso de segmentación de los latidos (Figura 5.7), se puede observar cómo el punto característico correspondiente a la onda R se detecta correctamente en prácticamente todos los latidos, consiguiendo de esta manera una perfecta segmentación del ECG.

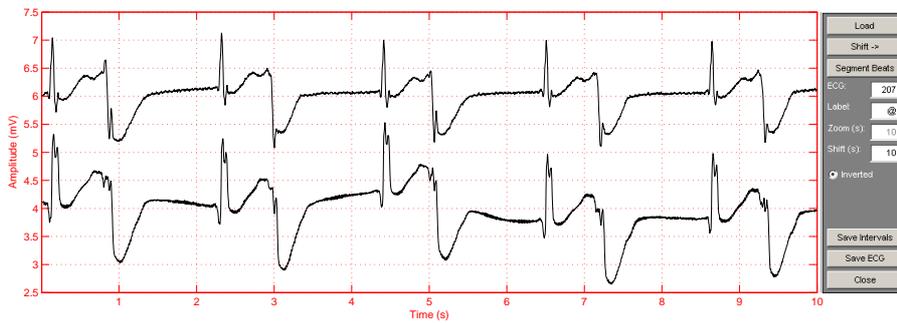


Figura 5.6: Aplicación desarrollada para proceder al etiquetado manual de los latidos.

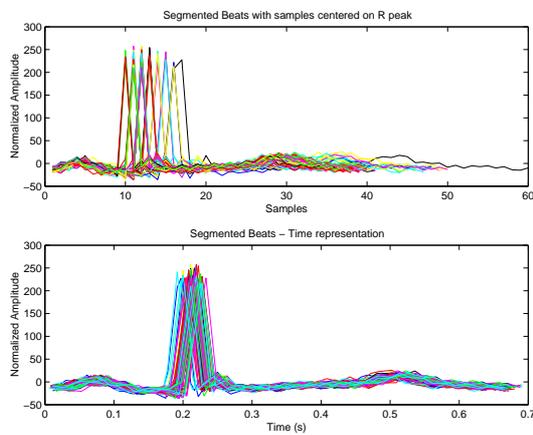


Figura 5.7: Segmentación de latidos en el ECG 205 del mitdb. (a) Representados por segmentos sin tener en cuenta su alineamiento temporal. (b) Alineadas temporalmente.

Por otra parte, podemos conseguir una importante reducción del tiempo de cálculo en el proceso de detección de los puntos característicos del latido y posterior segmentación de la señal Holter si, al igual que en el caso anterior, minimizamos el volumen de información a tratar, procesando la aproximación poligonal en lugar de las muestras originales. En un experimento de segmentación realizado sobre una señal de 89000 muestras, el procesamiento de la señal aproximada supone una mejora de hasta un 65 % sobre el tiempo

de procesamiento del ECG original.

5.6. Normalización de los latidos

La normalización de la señal consiste en igualar el orden de magnitud de las muestras pero manteniendo la morfología de los latidos. Normalizando los valores de la amplitud (que dependen de factores tan dispares como la amplificación del Holter, la conductividad de los electrodos, la fisonomía del paciente, etc.), se pretende un posterior análisis de distintas señales electrocardiográficas procedentes de un mismo paciente, que han podido ser adquiridas incluso mediante equipamiento y dispositivos diferentes. Otro de los motivos para la normalización de los latidos del ECG es, en vistas a una posterior reducción de las características mediante aplicación y entrenamiento de HMMs, el de mantener en todo momento los valores de la señal dentro de ciertas cotas que aseguren una convergencia del algoritmo EM que se utilizará en la fase de entrenamiento de los modelos. Para ello se recalculan los vectores de amplitud y duración para que presenten una media nula y una desviación típica igual a la unidad (*Figura 5.8*).

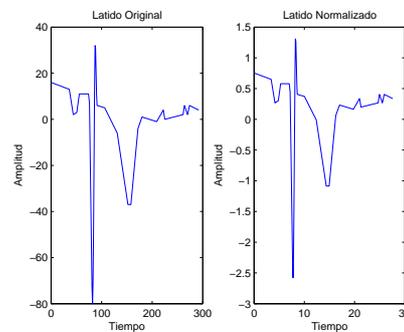


Figura 5.8: Normalización de los latidos donde se comprueba como, a pesar de modificar los vectores de amplitud y duración, se mantiene la forma de onda.

5.7. Conclusiones

Las conclusiones derivadas de los apartados anteriores son las que se describen a continuación:

- En el caso de las *interferencias introducidas por la red de alimentación* se filtra la señal de forma selectiva rechazando el contenido frecuencial localizado en la frecuencia que se desea eliminar.
- Para la eliminación de las *variaciones de la línea de base* se utilizan wavelets, y se comprueba como, su aplicación sobre la señal aproximada no supone una mejora sustancial (en cuanto a velocidad de procesamiento se refiere) respecto de la aplicación de las wavelets directamente sobre la señal original. Por este motivo la corrección de la línea de base se efectuará antes de la extracción de las características de la señal.
- Para la eliminación del *ruido interferente* se aprovecha el propio proceso de extracción de las características mediante aproximación poligonal. Además, las señales utilizadas en las pruebas presentan una muy buena SNR.
- La *segmentación* del ECG en latidos se realiza sobre la aproximación poligonal, obteniendo hasta un 65 % de ahorro en el tiempo de procesamiento.

Parte III

Clustering de latidos

Capítulo 6

Introducción al clustering

***Resumen.** A continuación se introducen una serie de conceptos básicos relacionados con el clustering de objetos, presentando algunos de los algoritmos utilizados en capítulos posteriores para el proceso de clustering sobre latidos de señales ECG de larga duración. La selección sobre el conjunto de datos del número de clusters adecuado también resulta de especial importancia para la optimización del proceso global. Además, se plantea el problema de los outliers, latidos que por sus especiales características devalúan la calidad del proceso de clustering. Finalmente se definen las medidas de calidad que nos ayudarán a validar los resultados del clustering obtenidos. En el apartado en el que se presentan las conclusiones y, a partir de las herramientas presentadas, se propone el objetivo de diagnóstico final a alcanzar en el proceso de clustering.*

Contenidos

6.1. Fases del proceso	153
6.2. Selección de características	153
6.3. Medidas de disimilitud	154
6.4. Criterios de clustering	155
6.5. Algoritmos de clustering utilizados	155
6.5.1. Algoritmo k-medias	156
6.5.2. Algoritmo Max-Min	158
6.5.3. Clustering jerárquico binario	160
6.6. Selección del número óptimo de clusters	163
6.7. El problema de los outliers	165
6.8. Diagnóstico y validación de los clusters	167
6.9. Conclusiones	170

6.1. Fases del proceso

El análisis de los datos, particionándolos de manera no supervisada en una serie grupos o clusters, se denomina clustering. En cualquier proceso de clustering se distinguen funcionalmente las siguientes etapas (*Capítulo 3*):

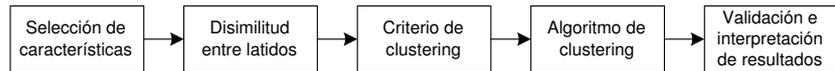


Figura 6.1: El proceso de clustering de latidos responde al diagrama presentado.

En el caso del clustering de latidos, el detalle de implementación de cada una de las etapas de la *Figura 6.1* es el que se presenta a continuación.

6.2. Selección de características

Para la etapa de selección de las características se parte de una serie de latidos previamente caracterizados y segmentados a partir de la señal ECG Holter tomada como fuente. Para alcanzar el éxito en el clustering posterior, resulta clave la selección precisa de las características anteriormente extraídas lo que supone que, de una manera ideal, se ha capturado únicamente la información crítica del objeto sin que la información redundante pueda interferir en su posterior procesado. Por todo esto y en los capítulos que continúan, se ha optado por comparar dos de los métodos aplicados en la etapa de selección de características de los latidos:

- (i) Métodos de análisis estadístico multivariante como el de *componentes principales o PCA* [AM99], descrito en la *Sección 3.4.1* del *Capítulo 3* y cuyos resultados se presentan en el *Capítulo 7*.
- (ii) Métodos de modelado como el de los *modelos ocultos de Markov o HMM* [Rab89], descrito en la *Sección 3.4.2* del *Capítulo 3* y cuyos resultados se presentan en el *Capítulo 8*.

6.3. Medidas de disimilitud

Tras la extracción de las características de los objetos a agrupar se procede al cálculo de la similitud (semejanza) o disimilitud (desigualdad) entre ellos [TK99], [Tan58], [CS96]. El conjunto de medidas entre objetos conforma una matriz de similitud (o disimilitud) cuyo procesamiento nos permitirá la posterior asignación de los elementos a uno u otro cluster dependiendo de la función criterio seleccionada. La formulación matemática de las funciones de similitud y disimilitud se puede encontrar en la *Sección 3.5.2 del Capítulo 3*. En el caso de medir la distancia de los latidos de una señal electrocardiográfica y para poder evaluar las prestaciones de cada uno de los métodos de clustering implementados, emplearemos las siguientes medidas:

- **Norma L_1 o City Block:** medida de disimilitud que, calculada sobre dos vectores x_i y x_j de una misma longitud, se define como:

$$L_1 = \sum_l \|x_i[l] - x_j[l]\| \quad (6.1)$$

La ponderación por un factor w_l de la norma L_1 es conocida como *distancia Manhattan*.

- **Norma L_2 o distancia euclídea:** medida de disimilitud que, utilizada en un espacio vectorial euclídeo entre los objetos x_i y x_j , se define:

$$L_2 = \sqrt{\sum_l (x_i[l] - x_j[l])^2} \quad (6.2)$$

- **Distancia euclídea normalizada:** definida del mismo modo que la distancia euclídea con el añadido de que, cada una de las coordenadas en la suma de cuadrados, resulta ponderada de forma inversamente proporcional a la varianza de las muestras de esa coordenada.
- **Distancia de Mahalanobis:** medida de disimilitud definida como sigue:

$$M = \sqrt{(x_i[l] - x_j[l]) \cdot V^{-1} \cdot (x_i[l] - x_j[l])'} \quad (6.3)$$

donde V es la matriz de covarianzas entre las variables de los elementos a medir.

- **Correlación:** medida de disimilitud definida como la unidad menos la correlación calculada entre los objetos a medir:

$$C = 1 - \frac{(x_i - \bar{x}_i) \cdot (x_j - \bar{x}_j)'}{[(x_i - \bar{x}_i) \cdot (x_i - \bar{x}_i)']^{\frac{1}{2}} \cdot [(x_j - \bar{x}_j) \cdot (x_j - \bar{x}_j)']^{\frac{1}{2}}} \quad (6.4)$$

- **Distancia de Hamming:** medida de disimilitud definida como el número de veces que dos vectores son diferentes [Lip87], [GG92]. En nuestro caso y al utilizar características de los latidos cuyos valores pertenecen al conjunto de los números reales \mathfrak{R} , esta medida resulta de poca aplicabilidad, por lo que no se utilizará en las pruebas.

6.4. Criterios de clustering

El criterio de clustering es el utilizado para evaluar la medida de similitud o disimilitud obtenida en el apartado anterior, decidiendo si un objeto pertenece o no a un determinado cluster y definiendo de esta manera la estructura de la agrupación final. El criterio de clustering se suele expresar mediante una función de coste. En muchos casos es la propia función de coste la que proporciona su nombre al algoritmo de clustering. Para los algoritmos seleccionados en el clustering de latidos de señales ECG Holter, los criterios de clustering utilizados son los que se presentan en la *Sección 6.5* donde se describe con detalle el funcionamiento de cada uno de ellos.

6.5. Algoritmos de clustering utilizados

Una vez adoptadas la medida de disimilitud y la función criterio, sólo nos queda seleccionar un algoritmo específico que revele finalmente la estructura interna del conjunto de datos, esto es, cómo aparecen realmente agrupados. A continuación se describen algunos de los algoritmos de clustering que utilizaremos en la clasificación de los latidos de la señal ECG Holter: el de *máxima*

similitud, el *k-medias*, el *Max-Min* y un algoritmo de clustering *jerárquico binario*. Para su descripción y formulación se utilizará la notación matemática introducida en el apartado dedicado a la *Formulación de los procesos de clustering* en la *Sección 3.5.2*.

6.5.1. Algoritmo k-medias

Se trata de un algoritmo basado en técnicas de agrupamiento particional ya tradicional en la literatura especializada [GT74], sobre el que se han propuesto numerosas variantes [Jua99] y que tiene como objetivo el minimizar la varianza intracluster maximizando la varianza intercluster. Básicamente se trata de minimizar un índice de distorsión definido como la suma de las distancias desde todos los objetos incluidos en un cluster a su centroide. El número de clusters k se determina a priori. El comportamiento del k-medias es sensible al número de clusters especificado, a la elección inicial de los centroides, al orden en que las muestras son tenidas en cuenta y, por supuesto, al conjunto de objetos en el espacio de representación. El algoritmo k-medias se resume en los siguientes pasos:

1. *Determinar los centroides iniciales* a partir de un conjunto X de objetos a agrupar y, bien conociendo a priori el número de clusters, bien estimándolo de alguna forma, se utilizan diversos métodos para escoger k objetos x_i para que éstos constituyan los centroides q_i de los correspondientes k clusters C_i :

$$C_1^1\{q_1^1\}, C_2^1\{q_2^1\}, \dots, C_k^1\{q_k^1\} \quad (6.5)$$

donde el superíndice indica la iteración en curso, inicialmente 1. Los criterios utilizados para la elección de los centroides son:

- *Aleatorio*, donde los centroides se eligen aleatoriamente de entre todos los objetos del conjunto de datos.
- *Uniforme*, donde los centroides se seleccionan uniformemente repartidos en el rango de datos.
- *Test*, realizando una prueba de clustering inicial sobre el 10% de los datos del conjunto para seleccionar los centroides resultantes

de dicho test inicial. A su vez, la selección de los centroides para este test resulta ser aleatoria.

2. **Distribución de los objetos** en los clusters, de tal forma que cada uno se asocia al cluster cuya disimilitud al centroide sea la mínima:

$$x \in C_j^t \text{ sii } \|x - q_j^t\| \leq \|x - q_i^t\|, \quad 1 \leq j, i \leq k, \quad i \neq j \quad (6.6)$$

donde x es el objeto genérico a clasificar, i y j son los índices de los clusters y t representa un índice de iteración genérico.

3. **Actualización de los centroides**, minimizando el índice de distorsión (Ecuación 6.7).

$$\gamma_i = \sum_{\forall x \in C_i^t} \|x - q_i^t\|^2, \quad 1 \leq i \leq k \quad (6.7)$$

Dependiendo de la métrica utilizada, el cálculo de los nuevos centroides resulta de una u otra manera. Las distancias utilizadas son:

- **Euclídea**, donde la elección de cada nuevo centroide se calcula como el vector medio (media muestral) de los elementos pertenecientes a dicho cluster C_i^t , de la forma:

$$m_i^{t+1} = q_i^{t+1} = \frac{1}{n_i^t} \sum_{\forall x \in C_i^t} x, \quad 1 \leq i \leq k \quad (6.8)$$

siendo n_i^t el número de elementos del cluster C_i en la iteración t .

- **Cityblock**, donde la distancia implementada es la suma de las diferencias absolutas y cada nuevo centroide es la media por componentes de las distancias de todos los puntos del cluster.
- **Coseno**, donde la distancia se calcula como la unidad menos el coseno del ángulo entre los objetos (considerados como vectores). El nuevo centroide se calcula como la media de los objetos normalizados del cluster.
- **Correlación**, donde la distancia se calcula como la unidad menos la correlación entre los objetos (considerados como secuencias de valores). El nuevo centroide se calcula como la media por componentes de los objetos centrados del cluster (media igual a cero y desviación estándar igual a la unidad).

4. *Fin del algoritmo* donde el agrupamiento termina cuando se alcanza la condición de estabilidad dada por la igualdad $q_i^{t+1} = q_i^t, \quad \forall i$. Mientras no se cumpla dicha condición, se repite el proceso desde la redistribución del resto de objetos (*paso 2*). Lo más normal en nuestro caso, en lugar de iterar hasta que se cumpla la condición de estabilidad, será el fijar un número máximo de repeticiones. Esto es debido a la dificultad intrínseca que presentan las señales ECG para minimizar su índice de distorsión, ya que no siempre se pueden representar de manera adecuada en un espacio euclídeo.

6.5.2. Algoritmo Max-Min

Como el algoritmo anterior, el Max-Min está basado en el agrupamiento particional (no jerárquico y no paramétrico) y puede considerarse como una aproximación al algoritmo voraz cuyo coste computacional es mucho más reducido [Jua99]. Como en este algoritmo el número de clusters no está definido a priori, será necesario el determinar un factor A , responsable de la creación o no de un nuevo cluster que se adapte a la estructura de los datos. Existen otras variantes del Max-Min donde el número de clusters sí que se define a priori. Este algoritmo se define como sigue:

1. *Definición del centroide inicial* a partir del conjunto de n puntos a agrupar X , en primer lugar se escoge aleatoriamente uno de los objetos disponibles x_i , y se crea un primer cluster C_1 , siendo el centroide $q_1 = x_i$. El resto de centroides se elegirá siguiendo una estrategia voraz donde el prototipo seleccionado en la iteración t es aquel cuyo prototipo más cercano entre los $t - 1$ ya existentes se encuentra más alejado.
2. *Definición del segundo centroide* calculando las disimilitudes de los $X - 1$ puntos restantes a x_i , y tomando aquel objeto que presente una disimilitud máxima x_j , el cual produce un nuevo cluster C_2 , con el centroide $q_2 = x_j$, donde:

$$j = \operatorname{argmax}\{d(x_l, x_i)\}, \quad 1 \leq l \leq n, \quad l \neq i \quad (6.9)$$

3. *Agrupación de los objetos* sobre los t clusters existentes hasta la iteración actual, asociando cada uno de los $X - t$ objetos restantes a aquel cluster cuya disimilitud sea mínima (o cuya similitud sea máxima).
4. *Elección de un nuevo centroide* aplicando la función criterio:

$$f(x_i) = \max\{\min\{d(x_i, q_j)\}\}, \quad x_i \in X - Q, \quad q_j \in Q \quad (6.10)$$

donde Q es el conjunto de centroides seleccionados. En este caso la función criterio compara las distancias entre los elementos y sus centroides y selecciona como nuevo centroide a aquel elemento cuya distancia a su centroide es la máxima de entre todos los objetos evaluados. Si el resultado supera cierto umbral (*Ecuación 6.11*), se considera el objeto x_i como centroide de un nuevo cluster. El proceso se repite desde el *paso 3* hasta que el resultado de la función criterio no supere el umbral δ .

$$\delta = A \cdot d(q_1, q_2) \quad (6.11)$$

donde A es una constante definida experimentalmente.

5. *Fin del algoritmo* ya que, una vez determinados los centroides de los clusters se agrupa cada objeto sobre el cluster cuyo centroide presente una menor disimilitud (*paso 3*). Tal y como se ha comentado, el proceso termina cuando todos los objetos están asignados a uno u otro cluster y ninguna de las disimilitudes intracluster supera el umbral especificado.

Un procedimiento alternativo consiste en fijar el número de clusters a utilizar, de manera que se salva el inconveniente de fijar un umbral, aunque en este caso no se podrá asegurar que la disimilitud intracluster esté por debajo de cierto valor. El nuevo centroide se toma como aquel objeto que maximiza la función criterio hasta que se completa el conjunto de centroides Q . Una mejora de este algoritmo incorporaría el criterio de estimación del número óptimo de clusters del conjunto de datos desarrollado en la *Sección 6.6*.

Una aplicación práctica del Max-Min se proporciona en el caso concreto de reducción de características en señales Holter mediante aplicación de HMM (*Capítulo 8*). En este caso, y al trabajar sobre una matriz de similitud, los objetos se clasifican en tantos clusters como modelos de que se disponga. El objeto

se asigna a aquel cluster cuyo HMM (centroide) proporcione una máxima probabilidad (similitud). En una nueva iteración del proceso se considera como nuevo centroide al objeto que presenta una menor probabilidad de ser generado por conjunto de centroides ya existente.

Finalmente cabría destacar que, como principal inconveniente, el Max-Min presenta el problema de una acertada selección de un valor umbral ya que, según sea su éste, la clasificación final puede variar de forma apreciable. Además, se trata de un algoritmo muy sensible a los *outliers* [Cue01].

6.5.3. Clustering jerárquico binario

Otro de los algoritmos que utilizaremos en el clustering de latidos está basado en técnicas de agrupamiento jerárquico, consistentes en la unión de los elementos de dos en dos (binario) para formar grupos mayores, resultando en un árbol de clusters denominado dendrograma, que muestra la relación entre los diferentes grupos de datos. Mediante el corte del dendrograma (Figura 6.2) en un cierto nivel se obtiene el clustering de los datos en grupos disjuntos (Capítulo 3, Sección 3.5.2). El algoritmo de clustering jerárquico binario se define en los siguientes pasos:

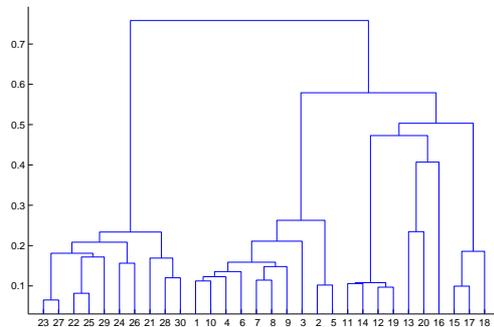


Figura 6.2: Dendrograma generado por la función criterio para el clustering binario de 30 objetos. Como se puede observar, en este caso se refleja la estructura interna de los datos, que revela la posible existencia de tres clusters diferentes, aunque el algoritmo de clustering todavía no los haya particionado.

1. **Determinar el cluster inicial** a partir de la matriz de disimilitud, el algoritmo debe agrupar los objetos en base a la función criterio determinada mediante la que se obtendrá inicialmente un cluster binario con dos latidos, que será considerado como un nuevo objeto del conjunto. Los criterios de clustering utilizados son los que se definen en el *paso 2*.
2. **Repetir el proceso de agrupamiento binario** descrito en el *paso 1*, aplicando el criterio de clustering seleccionado hasta que todos los objetos del conjunto hayan sido agregados a una estructura jerarquizada en forma de dendrograma. Los criterios de clustering jerárquico utilizados miden la proximidad entre dos grupos de objetos, agregando en el mismo cluster binario los dos objetos más próximos de entre todos los evaluados. Estos criterios son los que siguen a continuación, donde n_r representa el número de objetos existentes en el cluster C_r ; n_s es el número de objetos del cluster C_s , y x_{ri} es el i -ésimo objeto del cluster C_r :

- **Distancia más corta** o de asignación al *vecino más próximo*, donde se consideran como pertenecientes a un mismo cluster aquellos dos objetos (entendiendo por objeto tanto a los elementos iniciales a agrupar como cualquier cluster binario ya existente y creado en una iteración anterior del algoritmo) que presenten un mínimo en términos de distancia (menor parecido entre ellos si se ha utilizado una MD, y mayor semejanza si la medida utilizada ha sido la MS) (Ecuación 6.12).

$$d(r, s) = \min(d(x_{ri}, x_{sj})), \quad 1 \leq i \leq n_r, \quad 1 \leq j \leq n_s \quad (6.12)$$

- **Distancia máxima** o de asignación al *vecino más lejano* donde, para la construcción de la jerarquía se utiliza en este caso la mayor distancia existente entre los objetos de los dos grupos analizados (Ecuación 6.13), agregando en el mismo cluster a los grupos con menor distancia máxima.

$$d(r, s) = \max(d(x_{ri}, x_{sj})), \quad 1 \leq i \leq n_r, \quad 1 \leq j \leq n_s \quad (6.13)$$

- **Valor medio**, que utiliza como criterio de clustering el valor medio de las distancias calculadas entre todas las parejas de objetos de los

clusters r y s (Ecuación 6.14).

$$d(r, s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} d(x_{ri}, x_{sj}) \quad (6.14)$$

- **Centroide**, donde se utiliza como criterio para el agrupamiento la distancia entre el objeto y los centros de masas de los diferentes grupos sobre los que agrupar (Ecuación 6.15). En este caso la matriz de distancias sólo tiene sentido si se ha utilizado como medida de disimilitud la distancia euclídea.

$$\begin{aligned} d(r, s) &= d(\bar{x}_r, \bar{x}_s) \\ \bar{x}_r &= \frac{1}{n_r} \sum_{i=1}^{n_r} x_{ri} \\ \bar{x}_s &= \frac{1}{n_s} \sum_{j=1}^{n_s} x_{sj} \end{aligned} \quad (6.15)$$

- **Incremental**, donde la distancia criterio se obtiene mediante el incremento en la suma de los cuadrados de las distancias intracluster resultantes de la fusión de los grupos r y s (Ecuación 6.16, donde d_{rs}^2 es la distancia entre los clusters r y s definida en la Ecuación 6.15). La suma de los cuadrados de las distancias intracluster se define como la suma de los cuadrados de las distancias desde cada uno de los objetos del cluster a su centroide.

$$d(r, s) = \frac{n_r n_s d_{rs}^2}{n_r + n_s} \quad (6.16)$$

3. **Evaluar la formación de los clusters**, comparando la información proporcionada por las distancias entre objetos presentadas en el dendrograma, con la matriz de disimilitud original que resulta en una matriz de distancias *cophenéticas* cuya correlación constituye el *coeficiente de correlación cophenético*, donde un valor próximo a la unidad nos indica que la estructura jerárquica obtenida se ajusta a la realidad de los objetos evaluados. La distancia *cophenética* entre dos objetos se define como el nivel de proximidad en el que dichos objetos se agrupan en un mismo cluster por primera vez [TK99]. Este coeficiente se puede utilizar para comparar la calidad de los resultados obtenidos en los diferentes casos.

4. *Identificar los clusters de datos* decidiendo el nivel de corte del dendrograma (con lo que los datos aparecerán agrupados en determinado número de clusters), bien a partir de la inconsistencia entre sus nodos, bien definiendo el número de grupos que tratamos de indentificar (esto no sucede en nuestro caso ya que, a priori, no conocemos ese número):
 - Utilizando un *coeficiente de inconsistencia*, calculado para cada una de las ramas del dendrograma y cuyo valor se obtiene (asumiendo cierta profundidad de niveles) al comparar el nodo del árbol con el valor medio de los nodos existentes por debajo suyo un número de niveles igual a la profundidad requerida. A mayor valor del coeficiente, mayor es la inconsistencia del nodo calculado con el resto de datos. Valores de inconsistencia altos identifican grupos de datos candidatos a generar clusters. Una vez decidido el nivel de inconsistencia utilizado en el clustering, sólo se tiene que cortar el dendrograma por todos aquellos nodos cuyo coeficiente de inconsistencia con el resto de nodos sea superior al definido, lo cual resulta en el agrupamiento final de los datos.
 - Definiendo de forma arbitraria *el número de clusters* en los que queremos agrupar los datos, lo que resultará en la división de la jerarquía por el nivel en el que se corten tantas ramas como clusters se ha decidido crear. Como en el caso que nos ocupa, no conocemos el número de clusters a priori, este método de agrupamiento de los datos a partir de su dendrograma no se va a utilizar.
 - Utilizando cualquiera de los métodos que se describirán en la *Sección 6.6* dedicada al desarrollo de criterios para la selección del número óptimo de clusters de un conjunto de datos.

6.6. Selección del número óptimo de clusters

Una de las tareas más importantes en el proceso de clustering es la de identificar el número de agrupamientos óptimo en el que distribuir los datos. En algoritmos de tipo jerárquico una aproximación intuitiva al número de clusters óptimo se realiza utilizando el dendrograma e identificando aquellas agrupa-

ciones de datos con una vida más larga (*lifetime*). El parámetro de *lifetime* de un cluster se define como la diferencia entre el nivel jerárquico (o nivel de proximidad) en el que se crea el cluster y el nivel jerárquico en el que es absorbido en un cluster mayor [TK99]. Éste método presenta el inconveniente de depender de la subjetividad del observador que realiza el agrupamiento. Existen muchos otros métodos aplicables tanto a los algoritmos jerárquicos como a los paramétricos, que independizan el cálculo del número de clusters de la subjetividad del observador [DH73] [RJ74]. Una comparativa entre algunos de ellos es la que aparece en [GM85]. En nuestro caso y a partir del conjunto inicial de datos X , vamos a desarrollar dos de los métodos propuestos en [JT93].

1. **Método extrínseco.** Se requiere que el usuario determine cierto parámetro de medida. Esto implica la definición de una función $h(C_r)$ que mida la disimilitud entre los vectores pertenecientes a un mismo cluster C_r . Ejemplos de este tipo de funciones son:

$$\begin{aligned} h_1(C_r) &= \max\{d(x_{ri}, x_{rj})\}, \quad 1 \leq i, j \leq n_r \\ h_2(C_r) &= \text{med}\{d(x_{ri}, x_{rj})\}, \quad 1 \leq i, j \leq n_r \\ h_3(C_r) &= \sum_i \sum_j \frac{d(x_{ri}, x_{rj})}{2n_{C_r}}, \quad 1 \leq i, j \leq n_r \end{aligned} \quad (6.17)$$

donde d es la métrica utilizada y n_{C_r} es la cardinalidad del cluster C_r . Si definimos acertadamente un umbral θ para la función genérica $h(C)$, el algoritmo de clustering finaliza en el cluster \mathfrak{R}_t si:

$$\exists C_s \in \mathfrak{R}_{t+1} : h(C_s) > \theta \quad (6.18)$$

es decir, \mathfrak{R}_t es el final del clustering si existe un cluster C_s en \mathfrak{R}_{t+1} cuya disimilitud entre sus vectores $h(C_s)$ es mayor que el umbral θ .

Normalmente se define el umbral θ como:

$$\theta = \mu + \lambda\sigma \quad (6.19)$$

donde μ es el valor medio de la distancia entre cualquier par de vectores de X y σ es su varianza, mientras que λ queda definido por el usuario.

Esto significa que la responsabilidad de la definición de un buen valor umbral θ recae ahora sobre el parámetro λ , aunque la ventaja que supone el planteamiento de la *Ecuación 6.19* es que el nuevo parámetro λ puede ser estimado más fácilmente que θ .

2. **Método intrínseco.** En este caso, únicamente se tiene en cuenta la estructura del conjunto de datos X . De esta manera, el final del proceso de clustering se establece cuando se cumple la ecuación:

$$d_{min}^{SS}(C_r, C_s) > \max\{h(C_r), h(C_s)\}, \quad \forall C_r, C_s \in \mathfrak{R}_t \quad (6.20)$$

con,

$$d_{min}^{SS}(C_r, C_s) = \min\{d(x_{ri}, x_{rj})\}, \quad 1 \leq i \leq n_r, \quad 1 \leq j \leq n_s \quad (6.21)$$

En otras palabras, se termina el clustering cuando la disimilitud entre todas las parejas de clusters es mayor que la disimilitud interna (*self-similarity*) de cada uno de ellos.

Finalmente cabe comentar que los métodos descritos se basan en argumentos heurísticos y que sólo proporcionan un indicativo del número óptimo de clusters del proceso.

6.7. El problema de los outliers

Un outlier es un objeto a clasificar que, por su especial morfología, no se debería encontrar ubicado en ningún cluster sino que aparece aislado en el espacio de representación. En el caso de señales ECG, un latido se convierte en outlier cuando su forma de onda se corrompe variando extremadamente debido a señales interferentes, ruido, etc. Los outliers suponen un problema para los algoritmos de clustering ya que, en general, son bastante sensibles a este tipo de objetos, tanto a la hora de identificar el número de clusters de la estructura de datos como a la hora de clasificar los latidos (*Figura 6.3, Figura 6.4*

y Figura 6.5). Por ejemplo, hay que tener en cuenta que en muchos casos y dependiendo del algoritmo utilizado, un outlier puede convertirse en centroide de un nuevo cluster o, en el mejor de los casos, puede desplazar el centroide de un cluster existente de tal forma que se agrupen bajo una misma etiqueta objetos que realmente pertenecen a grupos diferentes. En los Cuadros 2.7 y 2.8 del Capítulo 2 se relacionan (identificados bajo la etiqueta 99) todos los outliers que aparecen en cada una de las señales ECG tratadas.

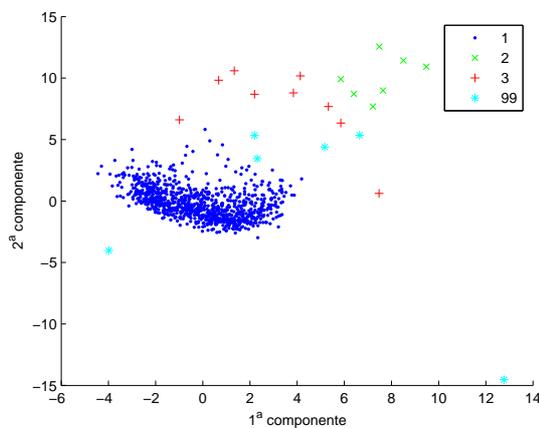


Figura 6.3: Representación bidimensional de una señal ECG caracterizada con PCA y formada por 953 latidos de tres clases diferentes y donde, además, aparecen 6 outliers (etiquetados con 99).

La solución al problema del clustering con outliers pasa por preprocesar los objetos en una etapa anterior a partir de la cual se identifiquen el mayor número de outliers de la manera más precisa posible. Al eliminar los outliers se podrá tratar a posteriori la señal con un algoritmo que trabaje sobre datos *limpios*, optimizando de esta forma los resultados obtenidos. En la Figura 6.5 se presentan las formas de onda de los latidos representativos de cada una de las clases tratadas en el ejemplo anterior. Etiquetado en la Clase 99 aparece la forma de onda de uno de los outliers generados en la etapa de segmentación de la señal ECG (no se han separado correctamente dos de los latidos).

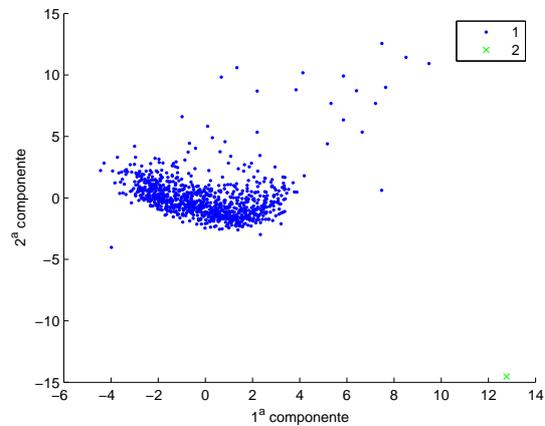


Figura 6.4: Clustering de la señal anterior (Figura 6.3) donde uno de los outliers dificulta el correcto agrupamiento de los objetos.

6.8. Diagnóstico y validación de los clusters

El reconocimiento de patrones y la toma de decisiones en tareas de clasificación presentan una serie de implicaciones que van más allá de la simple evaluación de una medida estadística. Para una correcta interpretación de los resultados será necesario contar con un criterio de diagnóstico, que nos ayude en esta toma de decisiones [Ran02]. En nuestro caso el diagnóstico es el proporcionado por los especialistas encargados de etiquetar convenientemente cada uno de los latidos de las señales electrocardiográficas.

Otro aspecto a tener en cuenta es el de la evaluación cuantitativa de los resultados obtenidos de la aplicación de un algoritmo de clustering concreto, proceso conocido como validación de los clusters [TK99]. Son muchos los métodos y herramientas desarrollados en este sentido y, en nuestro caso concreto, al contar con un diagnóstico a priori de los latidos el proceso de validación de los clusters obtenidos resulta relativamente sencillo. A partir de una secuencia de latidos (que presenta un número de X objetos a clasificar) y que son agrupados en una serie de clusters finales, donde x_r representa a un latido que pertenece realmente a la clase C_r y n_r es el número real de objetos existentes en esa clase, siendo T_i^r el hecho de clasificar el latido x_i en el cluster C_r . A par-

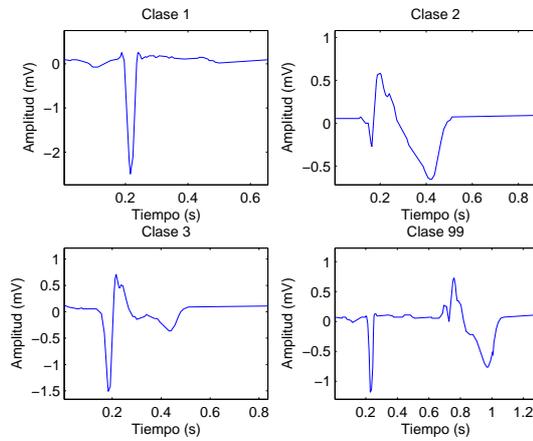


Figura 6.5: Formas de onda de los latidos representativos de cada una de las clases de la señal anterior (Figura 6.3).

tir de estas definiciones la validación de los resultados se calcula en términos de precisión, de la siguiente manera:

- $P(x_r)$: representa la probabilidad de los latidos pertenecientes a la clase C_r , donde $P(x_r) = \frac{n_r}{X}$.
- $P(x_{\bar{r}})$: representa la probabilidad de los latidos que no pertenecen a la clase C_r , donde $P(x_{\bar{r}}) = 1 - \frac{n_r}{X}$.
- $P(T_i^r)$: una vez terminado el clustering, es la probabilidad de haber clasificado un latido x_i en el cluster C_r .
- $P(T_i^{\bar{r}})$: una vez terminado el clustering, es la probabilidad de haber clasificado un latido x_i en un cluster distinto del C_r .
- TP (*true positive*): representa el hecho de clasificar un latido en el cluster correcto (también conocido como acierto). La sensibilidad S^+ o TPF (*true-positive fraction*) se define como: $S^+ = \frac{\sum TP}{n_r}$.
- TN (*true negative*): representa el hecho de rechazar desde un cluster un latido que está etiquetado como no perteneciente a ese cluster. La especificidad S^- o TNF (*true-negative fraction*) se define como: $S^- = \frac{\sum TN}{X - n_r}$.

- *FP (false positive)*: representa el hecho de clasificar en el cluster analizado un latido que está etiquetado como no perteneciente a ese cluster (fallo). La probabilidad de este hecho, conocida como *false-positive fraction* es: $FPF = \frac{\sum FP}{X - n_r}$.
- *FN (false negative)*: representa el hecho de rechazar en un determinado cluster un latido que realmente sí pertenece a ese cluster. La probabilidad de este error de clasificación, conocida como *false-negative fraction* es: $FNF = \frac{\sum FN}{n_r}$.

En este punto resulta necesario el recordar que, aunque el proceso de clasificación de latidos es no supervisado realmente y, al disponer de las señales convenientemente etiquetadas por el MIT [GAG⁺00], sí que podremos validar los métodos desarrollados mediante el cálculo de parámetros tales como su sensibilidad o especificidad. A partir de las definiciones anteriores se establecen las siguientes relaciones:

$$\begin{aligned}
 TPF + FNF &= 1 \\
 TNF + FPF &= 1 \\
 S^- = 1 - FPF &= TNF \\
 S^+ = 1 - FNF &= TPF
 \end{aligned} \tag{6.22}$$

Una medida que nos permite evaluar la calidad del cluster C_r , en términos de precisión (ACC_r) es la definida en la *Ecuación 6.23* donde se premian los aciertos y se penalizan los fallos ponderando el marcador final sobre el número total de objetos a clasificar.

$$ACC_r = \frac{\sum TP - \sum FP + \sum TN - \sum FN}{X} \tag{6.23}$$

En nuestro caso, definiremos la precisión total ACC_{total} del proceso de clustering para determinado ECG donde finalmente se han obtenido k agrupaciones como la suma ponderada de la precisión calculada anteriormente para cada uno de los clusters:

$$ACC_{total} = \sum_{i=1}^k ACC_i \cdot P(x_i) \quad (6.24)$$

donde valores de ACC_{total} próximos a la unidad indican una buena precisión en el proceso de clustering ejecutado.

6.9. Conclusiones

En este apartado se han presentado y desarrollado algunos de los diferentes criterios y algoritmos cuyas prestaciones serán evaluadas en capítulos posteriores mediante su aplicación práctica al clustering de latidos de señales de tipo Holter. Además de la importancia que, en el proceso global, suponen tanto la adecuada extracción de las características de los elementos a agrupar como la cuidada selección de la función de coste o criterio de clustering, también deberemos tener en consideración el problema añadido de los outliers, cuyo adecuado tratamiento puede suponer una mejora importante de los resultados finales del clustering. Por otra parte, también se han planteado las bases para la identificación y selección del número óptimo de clusters en que se distribuyen los datos analizados. Finalmente se ha propuesto una medida de precisión que utilizaremos para evaluar la calidad global de los resultados obtenidos en el proceso de clasificación no supervisada.

Capítulo 7

Clustering mediante Análisis de Componentes Principales

Resumen. En la etapa de reducción o selección de características del proceso de clustering se pueden utilizar una gran variedad de métodos entre los que destacan los de análisis estadístico multivariante. En este capítulo nos vamos a centrar en el análisis de las componentes principales de los latidos (PCA) para, a partir de las características obtenidas, agruparlos en clusters para su posterior análisis y diagnóstico. Los objetivos perseguidos en este capítulo son el de (i) la optimización de los resultados de clustering utilizando como método de reducción de características el PCA y el de (ii) la estimación de las prestaciones ofrecidas por los distintos algoritmos de clustering implementados en este caso. Para la validación de los resultados se utilizarán las medidas de precisión definidas en la Sección 6.8 del Capítulo 6.

Contenidos

7.1. Generalidades sobre las pruebas con PCA	173
7.2. Generación de la matriz de datos	173
7.3. Identificación de los outliers	175
7.4. Optimización del número de clusters	177
7.5. Optimización para el clustering K-medias	178
7.6. Optimización para el clustering Max-Min	179
7.7. Optimización para el clustering binario	180
7.8. Comparativa entre los algoritmos de clustering utilizados	181
7.9. Conclusiones	182

7.1. Generalidades sobre las pruebas con PCA

Para las pruebas de clustering que siguen y, en la etapa de selección de características, utilizaremos la técnica de análisis estadístico multivariante denominada PCA, considerando cada uno de los latidos de la señal ECG como un proceso estocástico definido por un número determinado de variables aleatorias. Para ello tomaremos como fuente la señal ECG segmentada en L latidos y aproximada poligonalmente.

Para el ajuste de los algoritmos de clustering se ha preparado un conjunto de experimentos donde, en cada uno de ellos, se optimiza uno de los parámetros tratados fijando para ello el valor del resto de los factores que caracterizan la prueba. Se han utilizado como fuentes un grupo de 45 señales ECG provenientes de la base de datos del MIT y que contienen un total de 44630 latidos. Finalmente, y al comparar los resultados obtenidos en los tests, se seleccionará el algoritmo de clustering óptimo en cada caso.

7.2. Generación de la matriz de datos

Para la generación de la matriz de datos será necesario el determinar los parámetros siguientes:

- *Propiedad utilizada*, ya que la aproximación poligonal de una señal electrocardiográfica genera un registro bidimensional con una componente temporal y otra que corresponde a la amplitud de la muestra. El problema se plantea en la unidimensionalidad del método de extracción de características seleccionado (PCA) que nos obliga a utilizar en la matriz de datos una de las dos componentes de la aproximación (amplitud o tiempo), o bien una combinación de ambas (pendiente, definida por el cociente amplitud/tiempo). Los resultados obtenidos en las pruebas realizadas para la selección de la propiedad óptima son los presentados en la *Figura 7.1*.
- *Número de variables* o segmentos m , que debe ser idéntico para todos los latidos del ECG por lo que se implementa un algoritmo que, a partir de

la aproximación poligonal del latido mediante n características (segmentos), selecciona los m (con $m \leq n$) segmentos requeridos de manera que se minimice el error. En nuestro caso y una vez analizados el número de segmentos utilizados en cada uno de los latidos de la señal aproximada, el número de variables quedará determinado por el número mínimo de los segmentos por latido utilizados en el conjunto de latidos que componen el ECG.

Una vez seleccionado el número de variables (en cuanto a número de segmentos por latido se refiere), se obtiene una matriz de datos de dimensión $L \times n$ sobre la que aplicaremos el método de extracción de características seleccionado (PCA). Para la selección de la matriz de datos óptima se han realizado una serie de pruebas de clustering sobre la misma fuente donde el único parámetro variado ha sido el de la propiedad utilizada en cada caso (amplitud, tiempo o pendiente). La descripción de la prueba y el valor de los parámetros fijados han sido los siguientes:

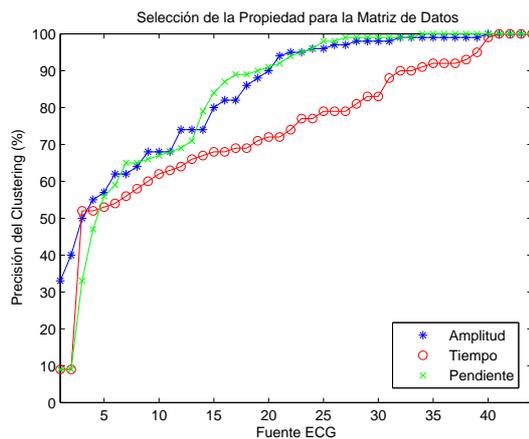


Figura 7.1: Resultados del clustering de latidos donde el parámetro variado ha sido el de la propiedad seleccionada en el proceso de extracción de características con PCA para la generación de la matriz de datos. El algoritmo utilizado es el Max-Min con un número de clusters exacto que ha sido definido a partir de los latidos etiquetados del MIT. Para una mayor claridad los resultados de los experimentos han sido ordenados de menor a mayor precisión.

7.3. Identificación de los outliers

Tal y como se ha comentado en la *Sección 6.7*, la optimización de los resultados del clustering pasa por la eliminación en la señal ECG del mayor número de outliers posible. El procesamiento previo de la señal durante la identificación de outliers debe ser lo suficientemente restrictivo como para que no se eliminen latidos reales cuyas especiales características puedan hacer que se confundan con éstos. La característica del algoritmo Max-Min de convertir en centroides de los nuevos grupos a aquellos objetos cuya disimilitud con el resto es la mayor, lo convierten en el candidato a ser utilizado en una etapa de clustering inicial que permita identificar los latidos aislados del conjunto, para ser etiquetados como outliers. En esta etapa de detección de outliers resultará primordial el minimizar el número de FP detectados (idealmente a cero) y no es tan importante el identificar un 100 % de TP ya que, para cualquier valor de TP superior a cero, estaremos eliminando outliers y, por lo tanto, mejorando los resultados del clustering posterior. Es decir, nos interesa maximizar el número de TP minimizando al mismo tiempo el número de FP.

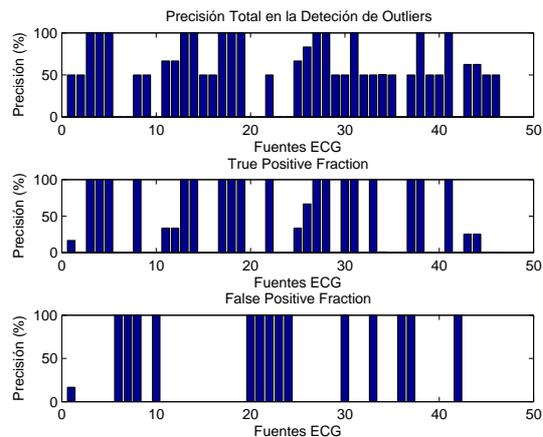


Figura 7.2: Precisión en el proceso de identificación de outliers (con asociatividad = 1) donde se ha utilizado el algoritmo Max-Min con detección automática del número de clusters. Los resultados se presentan mediante los estimadores de TPF, FPF y precisión para los outliers identificados.

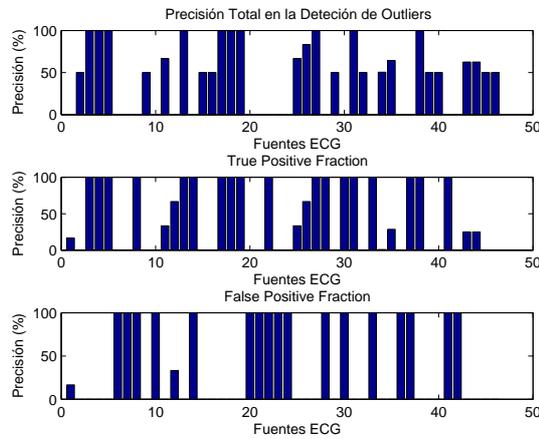


Figura 7.3: Precisión en el proceso de identificación de outliers (con asociatividad = 2) donde se ha utilizado el algoritmo Max-Min con detección automática del número de clusters. Los resultados se presentan mediante los estimadores de TPF, FPF y precisión para los outliers identificados.

En las pruebas de detección de los outliers realizadas y que han sido presentadas en las Figuras 7.2 y 7.3, el número de clusters de cada señal ha sido estimado por el algoritmo Max-Min de forma automática, implementando como criterio para establecer el final del proceso de clustering el método intrínseco propuesto en la Sección 6.6 del Capítulo 6. El parámetro definido como *Asociatividad* determina el número máximo de latidos que, asociados en un mismo grupo, podemos considerar como outliers. Los estimadores de calidad se calculan a partir de la comparación entre los resultados obtenidos en las pruebas y las etiquetas reales de los latidos del MIT. La precisión en la detección de los outliers se calcula como la resta entre el TPF y el FPF donde, para evitar la exclusión de objetos que no son outliers, si el índice FPF resulta mayor que cero, la precisión estimada es igual a cero (si se detecta como outlier un objeto que realmente no lo es se penaliza la precisión del método resultando nula).

7.4. Optimización del número de clusters

Las señales ECG originales utilizadas en las pruebas presentan un número de clusters (morfologías de los latidos distintas) determinado en las *Tablas 2.7* y *2.8* del *Capítulo 2* dedicado a la descripción de las señales ECG utilizadas. Para determinar la estructura óptima de los datos en cuanto a número de agrupaciones que estos presentan, se ha implementado un método de estimación intrínseco (donde el observador humano no interviene para definir ningún parámetro) que ha sido convenientemente detallado en la *Sección 6.6* del *Capítulo 6*. Los resultados obtenidos para el cálculo del número de clusters son los que se presentan a continuación:

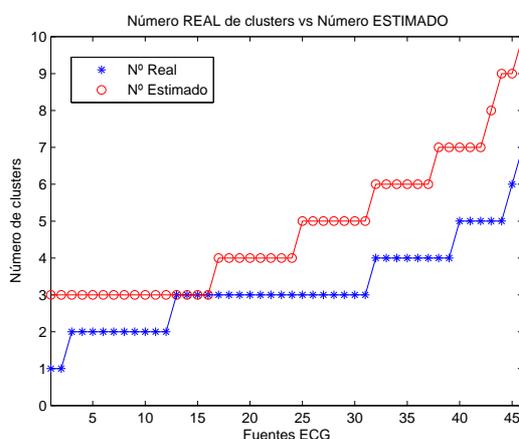


Figura 7.4: El criterio de selección intrínseco para la estimación del número óptimo de clusters implementado en el algoritmo de clustering Max-Min, con matriz de datos basada en la pendiente y para una extracción de características mediante PCA proporciona los resultados de la figura.

A partir de estos resultados (*Figura 7.4*) se puede confirmar que, efectivamente, los mejores resultados del clustering se obtienen cuando el número de grupos obtenidos coincide con el número de clusters que contiene la señal ECG y que es conocido a priori. En cuanto al método de estimación de la estructura interna de los datos (*Figura 7.5*), resulta demasiado optimista evaluando normalmente un número de grupos superior al que realmente existe.

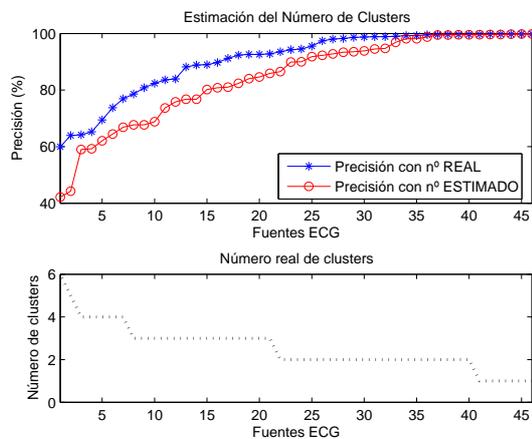


Figura 7.5: La precisión en el clustering depende del número de clusters utilizados en el algoritmo de clustering, alcanzando un óptimo cuando el número de clusters estimado coincide con el número de clusters real.

7.5. Optimización para el clustering K-medias

Tal y como se comentó en la *Sección 6.5*, el algoritmo K-medias es sensible al número de clusters especificado (y que necesariamente debe ser definido a priori), a la elección inicial de los centroides, al orden de las muestras y a la estructura de los datos. La fuerte dependencia que, sobre todos estos parámetros presenta el K-medias, lo convierten en un algoritmo de clustering la calidad de cuyos resultados varía en demasía por lo que será difícilmente aprovechable para nuestros propósitos. Además necesita de la definición a priori del número de clusters en que agrupar los datos, parámetro éste que se desconoce en un principio. De todas maneras, en el proceso de optimización del K-medias, y para poder establecer una comparativa con los resultados obtenidos mediante otros algoritmos de clustering, será necesario el probar con las diferentes métricas utilizadas en la minimización del índice de distorsión, seleccionando aquella que proporcione los mejores resultados.

En nuestro caso las métricas utilizadas son las que quedan reflejadas en la *Figura 7.6* donde, tal y como se aprecia, la más precisa resulta ser la distancia *Euclídea*. La elección de los centroides iniciales se ha efectuado a partir de una

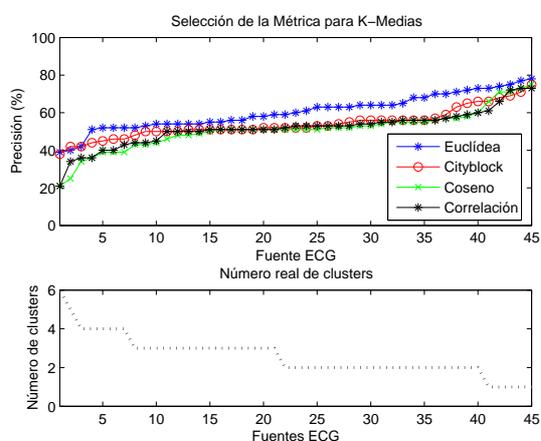


Figura 7.6: Resultados del clustering de latidos donde el parámetro variado ha sido el de la métrica utilizada para el algoritmo K-medias. La extracción de las características ha sido realizada por PCA manteniendo un 90% de la varianza de los datos. La matriz de datos incluye la pendiente de la aproximación poligonal.

prueba de clustering preliminar sobre un subconjunto compuesto por el 10% de las muestras a agrupar. A su vez, en la inicialización de estas pruebas preliminares los centroides se seleccionan aleatoriamente. El número de clusters definido en cada prueba ha sido el exacto (a partir de los latidos etiquetados, ver *Capítulo 2*). Además, cada prueba de clustering se ha repetido un total de cinco veces con un máximo de 1000 iteraciones en cada una de ellas.

7.6. Optimización para el clustering Max-Min

Para la optimización del algoritmo de clustering Max-Min será necesario el probar con las diferentes métricas que se utilizan durante el cálculo de las distancias entre los latidos, seleccionando aquella que proporcione los mejores resultados en cuanto a detección de agrupamientos se refiere. En nuestro caso las métricas utilizadas son las que aparecen en la *Figura 7.7* donde tal y como se observa, para un número de clusters inferior a 4, la que ofrece una mayor calidad en el clustering final es la distancia *Euclídea* mientras que para

un número de agrupamientos superior, la distancia *Cityblock* es la más precisa.

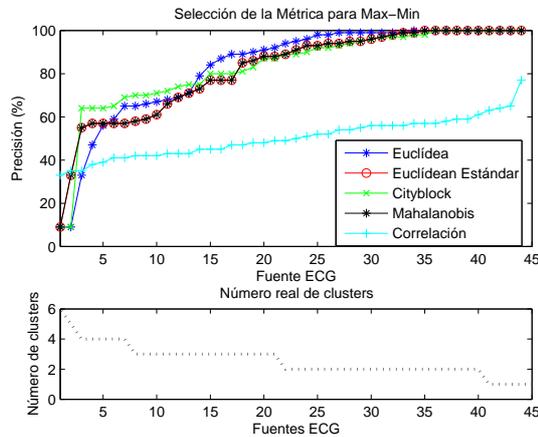


Figura 7.7: Resultados del clustering de latidos donde el parámetro variado ha sido el de la métrica utilizada para el algoritmo Max-Min. La extracción de las características ha sido realizada por PCA manteniendo un 90 % de la varianza de los datos. La matriz de datos incluye la pendiente de la aproximación poligonal.

7.7. Optimización para el clustering binario

Para la optimización del algoritmo de clustering Binario también probaremos con las diferentes métricas utilizadas durante el cálculo de las distancias entre los latidos, seleccionando aquella que proporcione los mejores resultados en cuanto a detección de agrupamientos se refiere. Curiosamente en este caso el algoritmo de clustering es independiente de la métrica utilizada, proporcionando los mismos resultados en cualquier caso.

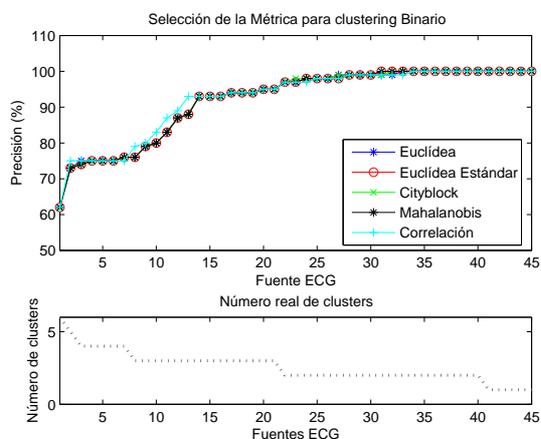


Figura 7.8: Resultados del clustering de latidos donde el parámetro variado ha sido el de la métrica utilizada para un algoritmo de tipo jerárquico binario. La extracción de las características ha sido realizada por PCA manteniendo un 90% de la varianza de los datos. La matriz de datos incluye la pendiente de la aproximación poligonal.

7.8. Comparativa entre los algoritmos de clustering utilizados

Para la comparación de los resultados obtenidos en la aplicación de los diferentes algoritmos de clustering utilizados se han seleccionado los resultados óptimos de cada una de las pruebas realizadas, generando la comparativa presentada en la *Figura 7.9* donde se aprecia como, de entrada, y sin perder de vista la especial estructura que presenta el conjunto de datos de una señal ECG, es el algoritmo K-medias el que peores resultados proporciona (tal y como cabía esperar debido a su fuerte dependencia de parámetros tales como la elección inicial de los centroides o su necesidad de conocer a priori el número de clusters en el que distribuir los datos). Por otro lado y para un número de clusters igual o superior a cuatro, es el algoritmo Binario el que optimiza ligeramente las prestaciones respecto al Max-Min mientras que, para un número de clusters inferior, los resultados ofrecidos por el Max-Min y el Binario son similares. Cuando el problema se complica al aumentar el número de clusters a detectar, la calidad de ambos algoritmos se iguala cayendo al 60%.

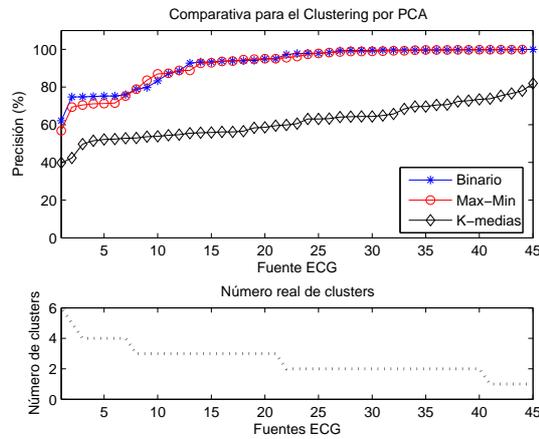


Figura 7.9: Resultados de la comparativa entre los diferentes algoritmos de clustering utilizados en el análisis de latidos de señales ECG Holter. La extracción de las características ha sido realizada por PCA manteniendo un 90 % de la varianza explicada. La matriz de datos incluye la pendiente de la aproximación poligonal.

7.9. Conclusiones

Las conclusiones obtenidas a partir de cada una de las pruebas son las referidas a continuación:

1. **Generación de la matriz de datos:** Las características de amplitud y pendiente son las que proporcionan unos mejores resultados de precisión en el clustering de latidos (*Figura 7.1*). En nuestro caso se utilizará la característica de pendiente en la que se incluyen tanto parámetros de amplitud como de tiempo (duración).
2. **Detección de los outliers:** Prueba en la que se utiliza el algoritmo Max-Min para tratar de detectar los outliers. Los mejores resultados en el proceso de detección se obtienen para un grado de asociatividad de 1 donde los outliers se identifican en la mayoría de fuentes con una precisión aceptable. Los casos en los que la precisión es nula (ha habido algún FP) se han producido cuando las fuentes no presentaban objetos etiquetados como outliers o cuando las características del objeto en cuestión,

aún no siendo un outlier, aparecen distantes a las del resto de objetos de su grupo. En este caso el latido puede ser eliminado sin problemas ya que, por las características de la señal ECG donde las morfologías de latidos se producen agrupadas en determinada secuencia temporal, el latido eliminado estará convenientemente representado en el clustering final por el resto de componentes de su grupo.

3. **Definición del número de clusters:** Se comprueba que los mejores resultados para el clustering de latidos se obtienen cuando el número de agrupaciones realizadas por el algoritmo coincide con el número de clusters reales (tipos de latido) existentes en la señal ECG previamente etiquetada. El método intrínseco utilizado para la evaluación del número de clusters resulta demasiado optimista detectando normalmente más clusters de los que realmente existen.
4. **Optimización de la métrica usando K-medias:** Independientemente del número de clusters, los mejores resultados los proporciona la distancia *Euclídea* (Figura 7.6). Comparando los resultados obtenidos con el número de agrupaciones reales que presentan las fuentes (línea discontinua), se observa como la calidad del clustering aumenta en tanto en cuanto el número de clusters a detectar es menor, es decir, a menor dificultad en el problema de clustering presentado, mejores son los resultados obtenidos.
5. **Optimización de la métrica usando Max-Min:** Para un número de clusters inferior a cuatro, los mejores resultados los proporciona la distancia *Euclídea* mientras que, en señales con mayor número de agrupamientos, la mejor métrica es la *Cityblock*. Tal y como ocurre en el caso anterior, la calidad del clustering también aumenta en tanto en cuanto el número de clusters a detectar es menor.
6. **Optimización de la métrica usando Binario:** En este caso los clusters generados por el algoritmo binario son independientes de la medida utilizada, proporcionando los mismos resultados en cualquier caso. Esto es debido a que los dendrogramas generados resultan similares con lo que el agrupamiento binario de los latidos es el mismo independientemente de la medida utilizada. Tal y como ocurre en los casos anteriores, aquí

también aumenta la calidad del clustering en tanto en cuanto el número de clusters a detectar es menor.

7. *Comparativa entre los algoritmos:* De la comparativa obtenida (Figura 7.9) se aprecia como son los algoritmos Max-Min y Binario los que ofrecen unas mejores prestaciones mientras que los resultados proporcionados por el algoritmo K-medias (debido a su fuerte dependencia respecto de la correcta selección de los centroides iniciales) son peores. También cabe resaltar que, conforme se complica el problema (al aumentar el número de clusters a detectar), las prestaciones de ambos algoritmos se igualan cayendo al 60%.

Si intentamos resumir los resultados obtenidos en este capítulo en una conclusión general podemos observar como, con la aparición de grandes clusters frente a clusters muy pequeños (y cuando el número total de clusters es reducido), se justifica el mal comportamiento de un algoritmo como el K-medias que trata siempre de *forzar* nuevas agrupaciones buscando precisamente sobre el cluster mayoritario, con lo que la calidad de los clusters obtenidos es pobre. En este caso y siempre que los clusters minoritarios aparezcan lo suficientemente alejados del mayoritario, un algoritmo particional como el Max-Min establecerá como centroides de los nuevos clusters a los elementos más distantes del centroide original, reagrupando los objetos a posteriori y mejorando los resultados del K-medias. Esta estrategia permite identificar los clusters pequeños con morfologías poco similares a la forma de onda de los latidos mayoritarios con lo que, en este caso, el Max-Min resulta el mejor de los algoritmos probados para aplicar en la tarea de detección de outliers. Lo mismo sucede con el clustering de tipo jerárquico binario, donde la inconsistencia entre pares de latidos nos lleva a agrupar los latidos morfológicamente similares, identificando los clusters minoritarios con un coeficiente de inconsistencia elevado respecto al cluster mayoritario. Es decir, el algoritmo K-medias funciona aceptablemente si la señal analizada no presenta outliers mientras que el Max-Min o el binario sí que detectan mejor los latidos de clusters minoritarios (outliers).

Finalmente cabría concluir que, independientemente del algoritmo de clustering utilizado, la métrica aplicada debería ser la distancia *Euclídea* con lo que, dependiendo de la dificultad del problema planteado (número de clus-

ters reales que presente la estructura de datos y aparición o no de outliers), la precisión en tareas de clustering simples puede superar el 90%.

Capítulo 8

Clustering utilizando Modelos Ocultos de Markov

Resumen. En este capítulo se presenta una herramienta que se va a utilizar tanto en la etapa de reducción de características de los diferentes tipos de latidos que aparecen en una señal electrocardiográfica como en el proceso de clustering propiamente dicho. Mediante los Modelos Ocultos de Markov (Hidden Markov Models, HMM) pretendemos modelar cada una de las morfologías para, a través de la probabilidad generada por el conjunto de modelos, conformar una matriz de similitud de reducidas dimensiones que facilite el posterior proceso de clustering sobre los distintos objetos. Tras una breve descripción teórica de los HMM y un enfoque sobre el problema del clustering de latidos utilizando estos modelos, el capítulo aproxima gradualmente la solución del clustering mediante el modelado de las diferentes clases. Además, en el apartado de pruebas y resultados se presentan todos los experimentos realizados. El capítulo termina con un apartado de conclusiones.

Contenidos

8.1. Introducción	189
8.2. El problema del clustering con HMM	190
8.3. Descripción del proceso de clustering con HMM	190
8.3.1. Adquisición y preprocesado	191
8.3.2. Selección de los centroides iniciales	193
8.3.3. Clustering de latidos	195
8.3.4. Entrenamiento	197
8.3.5. Consolidación y evaluación	199
8.4. Experimentos y resultados	201
8.4.1. Experimentos previos al clustering	201
8.4.2. Experimentos de clustering	209
8.5. Conclusiones	213

8.1. Introducción

En este capítulo se plantea el modelado estadístico como una de las técnicas a utilizar tanto para la reducción de las características como para la obtención de una matriz de similitud entre el conjunto de objetos a agrupar. El problema planteado en cualquier proceso de modelado es el de determinar el número de parámetros libres necesarios para diseñar un clasificador que proporcione buenos resultados. Este número debería ser el mínimo posible de tal manera que se maximice el ratio establecido entre los parámetros que determinan el clasificador l y el número de patrones de entrenamiento N de que se dispone, favoreciendo así las propiedades de generalización del clasificador resultante. Por otra parte, el hecho de representar los objetos mediante un elevado número de características se traduce en la necesidad de un gran número de parámetros sobre el clasificador. Como lo más normal es que se disponga de un número finito y bastante limitado de patrones de entrenamiento lo deseable sería la definición de un clasificador con el menor número de parámetros posibles, en aras de mejorar su capacidad de generalización sobre las secuencias de datos. Para conseguir el sistema clasificador de mejores prestaciones deberemos maximizar el ratio N/l , mejorando la estimación del error y consiguiendo de esta forma el objetivo de calidad perseguido. En la mayoría de los casos se aconsejan ratios del orden de 20 a 30 [TK99].

En cuanto al trabajo desarrollado en este capítulo se refiere, y siempre en la línea de optimización del clustering de latidos de señales ECG de tipo Holter, trataremos de reducir las características de los objetos utilizando para ello los HMMs. De hecho, el sucesivo modelado de los centroides obtenidos en cada iteración del proceso (ver *Figura 8.1*) y la posterior proyección de los latidos sobre el conjunto de modelos proporciona una matriz de similitud (la medida en este caso vendría dada en términos de probabilidad) sobre la que poder aplicar distintos algoritmos de clustering. El objetivo propuesto es el de maximizar la distancia *intercluster* minimizando al mismo tiempo la varianza *intracluster* utilizando los HMMs para la selección del menor número de características que se ajuste al objetivo citado. La meta final es siempre la de simplificar y mejorar la posterior aplicación de métodos de clustering estático sobre la matriz de similitud obtenida. A continuación se exponen los problemas planteados además de una descripción detallada del proceso.

8.2. El problema del clustering con HMM

Uno de los principales problemas que presenta el modelado de series temporales de datos es el de la adecuada selección de la estructura del modelo. En el caso de un HMM resulta muy importante tanto la configuración inicial del modelo (matriz de transiciones entre estados, probabilidades de emisión y probabilidades iniciales) como su dimensionamiento (utilización del número óptimo de estados). En [LB00] se presenta un algoritmo que trata de resolver ambos problemas de forma anidada en cuatro pasos (*Algoritmo 8.1*).

- 1) *Determinar el número de clusters de la partición*
 - 2) *Agrupar los objetos en los clusters*
 - 3) *Determinar el número de estados del HMM para cada cluster*
 - 4) *Actualizar el HMM para cada cluster*
- (8.1)

En nuestro caso, proponemos una serie de mejoras sobre este algoritmo donde, en el *Paso 1* se implementa el método intrínseco para la estimación del número de clusters propuesto en la *Sección 6.6*. Además, en el *Paso 3* se comprobará cuál de las estrategias de dimensionamiento del HMM propuestas en [BMF03] o [LB00] proporciona mejores resultados. El diagrama de bloques utilizado en este caso es el que se presenta en la *Figura 8.1*. Finalmente también se puede añadir una etapa de consolidación de modelos en la que se evalúan las prestaciones de los distintos HMMs obtenidos para, si procede, agrupar objetos sobre un mismo cluster.

8.3. Descripción del proceso de clustering con HMM

En las pruebas de clustering que siguen se utilizan los *Modelos Ocultos de Markov*, tomando como fuente de datos la señal ECG segmentada en latidos y aproximada poligonalmente. Tal y como ya se ha comentado, lo que se pretende mediante la utilización de los HMM es el seleccionar las características de los objetos a agrupar para, de esta manera, simplificar al máximo el proceso de clasificación. Aprovechando la capacidad de generalización de los HMMs y utilizando para su entrenamiento las secuencias de latidos pertenecientes

a una misma clase, trataremos de optimizar la caracterización de los objetos maximizando la distancia *intercluster* y reduciendo al mismo tiempo la varianza *intracluster*. El proceso propuesto para el clustering es el que se muestra en la *Figura 8.1* donde se repiten las iteraciones (refinando los modelos mediante entrenamiento) hasta que se cumpla el criterio de clustering seleccionado. Tal y como se puede apreciar en la etapa de selección de los centroides iniciales, realmente en ésta se implementa el algoritmo Max-Min. La idea final del proceso sería la de caracterizar los latidos con tantos modelos como clusters se identifiquen. A continuación pasaremos a describir con detalle cada una de las etapas y todos los pasos en que se ha dividido el proceso completo.

8.3.1. Adquisición y preprocesado

En esta *Etapa 1* se pretenden aglutinar todas las operaciones encaminadas a la obtención, preparación, caracterización y segmentación de los objetos y que ya han sido tratadas con detalle en otros capítulos de esta tesis (*Parte II*). En concreto los pasos desarrollados han sido los siguientes:

Adquisición de la señal

Realizada mediante algún dispositivo de tipo Holter ya que se trata de señales ECG de larga duración. Parámetros relativos al proceso de adquisición tales como la frecuencia de muestreo o la resolución (bits por muestra) vendrán determinados por las características del dispositivo de adquisición. Detalles como el de la composición en latidos de las señales utilizadas en nuestro caso se pueden consultar en el *Capítulo 2*.

Filtrado del ruido

Incluyendo en esta etapa la eliminación de la interferencia de la red y la corrección de la línea de base (*Capítulo 5*).

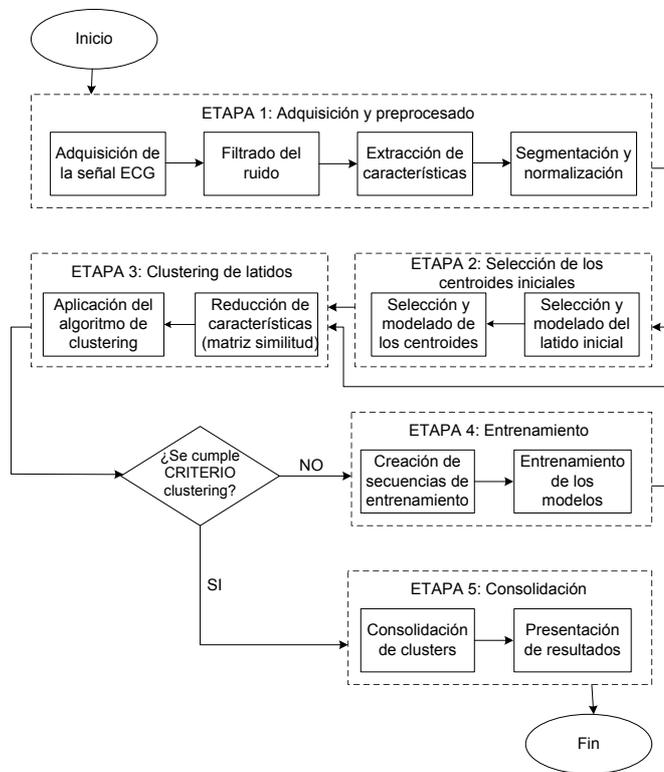


Figura 8.1: Diagrama de bloques propuesto para el clustering de señales Holter ECG mediante modelos ocultos de Markov.

Extracción de características

Para lo que se utiliza una aproximación poligonal de la señal ECG que, de esta forma, se convierte en el paso más importante de esta primera etapa. La extracción de las características supone la transformación de una larga serie de muestras de amplitud equidistantes en el tiempo (distancia temporal determi-

nada por la frecuencia de muestreo del sistema de adquisición) en una serie mucho menor (dependiendo del ratio de compresión utilizado) de muestras bidimensionales (amplitud, tiempo) cuya unión mediante segmentos significará una aproximación de la señal original sin pérdida de información relevante. De esta manera se consigue no sólo la reducción del tamaño de fichero a tratar sino que, además, se elimina parte del ruido interferente. Si, en cálculos posteriores como el de segmentación, se actúa sobre la señal comprimida, se consigue una considerable reducción del tiempo de procesamiento.

Segmentación y normalización

En este caso se analiza la señal ECG para identificar y separar los latidos que la componen. Una vez segmentada la señal, y con el fin de conservar la forma de onda de cada latido dentro de unos valores acotados, se procede a la normalización de los mismos. El escalado resulta adecuado en este caso ya que nos permitirá la comparación entre formas de onda adquiridas desde dispositivos diferentes, incluso proviniendo de pacientes diferentes (*Capítulo 5*). Hay que tener en cuenta que los valores de amplitud eléctrica pueden variar por causas tan diversas como el sudor, cambios en la conductividad de la piel o de los electrodos, la adquisición con dispositivos Holter distintos, etc.

8.3.2. Selección de los centroides iniciales

El problema que se plantea en esta *Etapa 2* es el de una adecuada selección inicial del primer latido que actúe como centroide para el clustering inicial que asegure posteriormente la utilización de secuencias de latidos de una misma clase en el entrenamiento de los modelos. Para ello se proponen las siguientes fases:

Selección y modelado del latido inicial

La selección del latido que utilizaremos para la inicialización del primer HMM se realiza de manera aleatoria sobre el total de latidos que componen la señal ECG. El modelo creado a partir del latido seleccionado nos permitirá

una reducción extrema de las características a una sola. Representando en una recta los valores de probabilidad de los latidos se puede asegurar que los objetos agrupados en los extremos de la recta pertenecen a clases diferentes. En esta aproximación inicial de un HMM a partir de las características del latido seleccionado se define un número máximo de estados del modelo para, a partir de éste, utilizar los diferentes métodos de estimación propuestos a continuación:

- **Aleatoria:** donde tanto la matriz de transición entre estados A , como la matriz de probabilidades de emisión B se proponen aleatoriamente [BMF03]. El vector de probabilidades iniciales de los estados π se fija a un valor de $[1, 0, \dots, 0]$ para forzar que la secuencia generada empiece siempre a partir del primero de los estados del modelo.
- **Aleatoria con máxima probabilidad:** donde se fija un número de iteraciones. En cada iteración se proponen aleatoriamente los parámetros del modelo y se calcula la probabilidad de que la secuencia de datos haya sido generada por dicho modelo. Al final del proceso se selecciona el conjunto de parámetros que han generado máxima probabilidad.
- **Viterbi:** donde, a partir de las matrices aleatorias A y B y teniendo en cuenta la secuencia de datos asociada al centroide a modelar, se utiliza el algoritmo de Viterbi para calcular la secuencia de estados más probable. A continuación se utiliza esta secuencia de estados para recalcularse la matriz de transición entre estados A . La matriz B se actualiza a las probabilidades de cada símbolo, calculadas a partir de la secuencia de datos.
- **Izquierda a derecha:** método de inicialización que se basa en la morfología del latido seleccionado como centroide para obtener una estimación ajustada de los parámetros del modelo [MCN04b].

Selección y modelado de los centroides iniciales

Como se ha comentado en el punto anterior, representando en una recta los valores de probabilidad (rigurosamente hablando lo que se representa es el valor del logaritmo de la probabilidad, a partir de ahora probabilidad

logarítmica) obtenidos a partir de la proyección de los latidos sobre el HMM inicial, se puede asegurar que los objetos agrupados en los extremos de la recta pertenecen a clases diferentes. Una vez elegidos los dos latidos que constituyen los centroides iniciales y utilizando alguno de los métodos de estimación anteriores, se calculará un HMM para cada uno de ellos [MCN04b], lo que nos permitirá obtener en el siguiente paso la matriz de similitud por proyección de los objetos sobre el modelo de cada centroide.

8.3.3. Clustering de latidos

En esta *Etapa 3* se proponen los HMMs como técnica para una reducción de las características de los objetos (obtención de un vector de probabilidad de longitud igual al número de centroides modelados, al proyectar el latido sobre cada uno de ellos) o, lo que sería equivalente en este caso, para la generación de la matriz de similitud entre objetos (con la i -ésima fila de la matriz correspondiendo con el reducido vector de características del i -ésimo objeto). La similitud entre un objeto y el centroide de un cluster se calcula a partir de la probabilidad de que el latido sea generado por dicho centroide, mientras que el criterio y el posterior algoritmo de clustering aplicados sobre la matriz de similitud dependen de las necesidades del usuario. Tras el agrupamiento de los latidos se verificará el cumplimiento de la condición de clustering donde, en caso afirmativo, se finalizará el proceso con una presentación de resultados. Si la condición no se cumple, se actualizarán los modelos en una *Etapa 4*.

Reducción de características

Dada la secuencia de puntos que representa un latido y el conjunto de modelos, el problema de la reducción de características (a tantas como de modelos se disponga) se limita al cálculo de la probabilidad de que dicha secuencia de puntos haya sido generada por cada uno de los modelos. Este cálculo se conoce con el nombre de problema de *inferencia* (ver *Sección 3.4.2*) donde dicha probabilidad se obtiene a partir del análisis conjunto de las probabilidades de la secuencia de datos sobre el HMM viajando tanto hacia adelante (forward) como hacia atrás (backward). Reduciendo las características de todos los objetos obtendremos una matriz de similitud de orden $T \times M$ (donde T representa

el número de objetos y M el número de modelos).

Algoritmo de clustering

A partir de la matriz de similitudes obtenida, se aplica el algoritmo de clustering seleccionado (ver *Capítulo 6*). Además de las nuevas etiquetas de los objetos agrupados, el algoritmo deberá proporcionar también la nueva secuencia de centroides a partir de la cual poder calcular los nuevos modelos en iteraciones posteriores.

Criterio de clustering

El criterio de clustering es el que determina si resulta necesario ejecutar una nueva iteración sobre las *Etapas 3 y 4* o se finaliza el proceso con la presentación de los resultados en una *Etapa 5*. Algunos de los criterios de clustering utilizados en los experimentos son:

- **Precisión de los resultados:** Para evaluar la precisión de los resultados del clustering se utilizan los estimadores detallados en el *Capítulo 6*. Mediante la comparación entre dos iteraciones consecutivas de estimadores tales como la precisión total, o el número de TP, FP, TN o FN, se puede establecer un criterio de finalización del clustering si, por ejemplo, el valor del estimador seleccionado no ha mejorado (o ha mejorado por debajo de cierto umbral establecido) respecto del valor del mismo estimador obtenido en la iteración previa. Este criterio lo podremos utilizar siempre y cuando, en un entorno de pruebas, se disponga de una secuencia de latidos etiquetados a priori.
- **Número de iteraciones:** Otra opción perfectamente válida como criterio de finalización del clustering es el de establecer un número máximo de iteraciones tras el cual se presenten los resultados. El número de iteraciones seleccionado deberá ser lo suficientemente elevado como para asegurar la convergencia del proceso.
- **Número de clusters:** Estableciendo como criterio de finalización el de un número de clusters detectados tras el cual se presenten los resultados. En

este caso se puede implementar cualquiera de los dos métodos citados en la *Sección 6.6* del *Capítulo 6* (intrínseco o extrínseco) para la detección automática del número de clusters.

8.3.4. Entrenamiento

A la *Etapa 4* de la *Figura 8.1* se llega tras no haber alcanzado el criterio de clustering establecido y en ella se procede a la actualización de los modelos utilizando como secuencia de entrenamiento parte del conjunto de los objetos asignados al cada uno de los clusters en la última iteración. Al mismo tiempo se da la posibilidad (a costa de ralentizar el proceso global) de optimizar el HMM mediante una selección del número óptimo de estados que lo componen. Para este dimensionamiento se utilizarán los algoritmos propuestos en [BMF03] y [LB00].

Creación de las secuencias de entrenamiento

La creación de la secuencia de entrenamiento para un modelo es un proceso delicado que requiere de la máxima precisión. Tal y como se ha comentado en la introducción al presente capítulo, la maximización del ratio N/l depende de una simplificación y un correcto dimensionamiento de los modelos. Para ello será necesario contar con un número suficiente de patrones de entrenamiento (latidos de una misma clase). Nuestro problema radica en asegurar que todos los latidos que se van a utilizar en el entrenamiento de determinado modelo presentan una misma morfología. En caso contrario (entrenando un modelo con objetos de varias clases diferentes) el modelo obtenido al final del proceso de entrenamiento no reflejaría adecuadamente las características de la clase (en términos de probabilidad), con lo que el cálculo de la matriz de disimilitud y la aplicación posterior del algoritmo de clustering derivaría en unas clases donde se confundirían latidos de morfologías diferentes. Para los experimentos realizados, el número de latidos a utilizar para el entrenamiento del modelo (secuencia de entrenamiento) se fija a priori seleccionando de un cluster aquellos objetos que presentan una probabilidad máxima sobre el modelo en cuestión (se trata de asegurar en la medida de lo posible que todos los latidos son de la misma morfología). El principal problema de esta etapa

se materializa en la falta o escasez de objetos asociados a un cluster con lo que la secuencia de entrenamiento resulta demasiado corta y por tanto el modelo obtenido no proporciona una probabilidad óptima (queda sub-entrenado).

Entrenamiento de los modelos

Los procesos de entrenamiento y optimización de los modelos se desarrollan paralelamente. El entrenamiento del modelo $\lambda = \{A, b, \pi\}$ consiste en el ajuste de sus parámetros para maximizar la probabilidad $P(\mathbf{O} \mid \lambda)$. Este problema se conoce con el nombre de *aprendizaje* (Capítulo 3.4.2). Debido a las especiales características de las series temporales de datos a clasificar, para el entrenamiento de los modelos se ha utilizado el algoritmo EM, conocido también como algoritmo de *Baum-Welch* (en su aplicación sobre los HMM). En las pruebas de clustering realizadas el número máximo de iteraciones se ha limitado a 200. Por otra parte está la optimización de un HMM que siempre pasa por su entrenamiento. El problema añadido cuando se pretenden modelar series temporales de datos es el del desconocimiento del número de estados necesario para que el modelo genere dichas series con un compromiso entre la simplicidad de su estructura y la probabilidad máxima obtenida. Nosotros hemos querido separar la fase de entrenamiento de la de selección del número de estados óptimo por lo que, aunque el problema de la selección del número de estados es conocido con el nombre de *selección* del modelo, nosotros lo denominaremos *optimización* del modelo. Para ello se proponen los siguientes métodos:

- **Método incremental:** [LB00] en el que se establece un criterio de evaluación (*Mixture minimum description length*, MMDL o *Bayesian inference criterion*, BIC) y número de estados mínimo k_{min} y máximo k_{max} . Partiendo de k_{min} y hasta k_{max} se van calculando los diferentes modelos a partir de las estimaciones iniciales obtenidas por cualquiera de los métodos presentados anteriormente. El modelo óptimo será aquel que maximice el criterio de evaluación seleccionado.
- **Método decremental:** o estrategia de poda (del inglés *pruning*) [BMF03] que consiste en el entrenamiento del modelo a partir de un número de estados máximo k_{max} estimado experimentalmente. El número de es-

tados k_{max} del modelo inicial $\lambda_k^I = \{A, B, \pi\}$ depende del método de estimación utilizado. La calidad del HMM resultante se calcula a partir del criterio seleccionado (MMDL o BIC). A continuación se reduce el número de estados obteniendo el modelo $\lambda_{k_{max}-1}$ de orden $k_{max} - 1$ a partir de la eliminación del estado menos probable del modelo de orden superior. El proceso de evaluación de los modelos se repite hasta alcanzar un número de estados mínimo k_{min} , también estimado de manera experimental. El modelo óptimo será aquel en el que se maximice el criterio de evaluación. El método descrito es el que se presenta en el *Algoritmo 8.2*.

- 1) Selección del criterio de evaluación (MMDL o BIC), k_{max} y k_{min}
- 2) Inicialización del modelo λ_k^I con k_{max} mediante alguno de los métodos propuestos
- 3) Mientras $k_{max} \geq k_{min}$ hacer:
 - a) Entrenar el modelo hasta alcanzar el criterio de convergencia, obteniendo $\hat{\lambda}_k$
 - b) Estimar el modelo guardando el valor criterio CR_k
 - c) Encontrar el estado menos probable (el menor elemento de p_∞), donde p_∞ es la distribución de probabilidad estacionaria de λ_k^I
 - d) Eliminar el estado menos probable obteniendo un modelo reducido $\bar{\lambda}$
 - e) Fijar $\lambda_{k-1}^I \leftarrow \bar{\lambda}$, $k \leftarrow k - 1$
- 4) Seleccionar el modelo con número óptimo de estados λ^* que maximiza CR_k :

$$\lambda^* = \hat{\lambda}_{k^*}, \quad \text{donde } k^* = \operatorname{argmax}_k (CR_k)$$

(8.2)

8.3.5. Consolidación y evaluación

En la *Etapa 5* de consolidación se pretende la compactación de los clusters fundiendo en un mismo grupo todos aquellos objetos pertenecientes a centroides similares. Tras las etapas de entrenamiento y optimización se evalúan los HMMs mediante una serie de medidas de calidad, comparándolos dos a dos para determinar si se trata de dos centroides que realmente pertenecen a un mismo cluster. Si las medidas de calidad de los dos centroides aparecen dentro de ciertos límites (establecidos experimentalmente) se seleccionará como centroide representativo del cluster al modelo que mejor se ajuste al

nuevo grupo de datos. Además también se añade un procedimiento para la presentación y evaluación de los resultados.

Consolidación de clusters

En esta etapa se evalúan los HMMs para que, en el caso de generar secuencias de datos parecidas, se consoliden en un solo cluster, agrupando todos los objetos sobre el centroide más simple (bien el HMM con un menor número de estados o bien el HMM que proporcione una mayor probabilidad sobre la secuencia de latidos). Para decidir si se consolidan dos modelos será necesaria la evaluación de la calidad de los HMMs que los representan. Dicha evaluación se realiza a partir del estimador de similitud intra-cluster (*Between Cluster Similarity*, BCS). Este estimador determina el grado de similitud existente entre dos modelos en términos de probabilidad. Para ello se comparan los centroides de los clusters dos a dos calculando la probabilidad de las series temporales de datos generadas por uno de los modelos dado el HMM del otro centroide y viceversa. Por ejemplo, si tenemos dos clusters C_i y C_j con modelos λ_i y λ_j respectivamente y series de datos asociadas a cada uno de ellos S_i y S_j , la BCS se calcula tal y como se propone en la Ecuación 8.3.

$$BCS(C_i, C_j) = \frac{\max_{S_i \in C_i} P(S_i | \lambda_j) + \max_{S_j \in C_j} P(S_j | \lambda_i)}{2} \quad (8.3)$$

Calculando la BCS para una partición con n clusters se obtiene una matriz de distancias o disimilitudes intra-cluster donde, los clusters susceptibles de ser agrupados son aquellos que, sin pertenecer a la diagonal, presentan el valor de BCS máximo.

Evaluación y presentación de resultados

Una vez finalizado el clustering sólo queda la evaluación y presentación de los resultados obtenidos para lo que se utilizarán los estimadores de calidad (precisión, sensibilidad y especificidad) detallados en el Capítulo 6.

8.4. Experimentos y resultados

Este apartado se divide en dos bloques. En el primero se efectúan una serie de pruebas previas al clustering de latidos, diseñadas con la intención de refinar en la medida de lo posible los distintos parámetros utilizados en cada uno de los algoritmos implementados (variando un parámetro y manteniendo el resto constante de manera que se puedan extraer conclusiones relativas al parámetro probado). En el segundo bloque de experimentos se presentan los resultados del clustering de latidos que componen una señal electrocardiográfica de tipo Holter. En todas las pruebas se han utilizado como fuente de datos las señales ECG de la base de datos del MIT [GAG⁺00] convenientemente preprocesadas y aproximadas poligonalmente con un total de 25 segmentos por latido. En el caso de los latidos de una señal ECG, queda adecuadamente justificada la utilización de objetos dimensionados con 25 características bidimensionales (duración, amplitud) ya que, con este número de segmentos se simplifica enormemente el volumen de información a tratar (un latido de duración 1s adquirido a 360Hz supone un total de 360 características de amplitud) sin perder en ningún momento las diferencias morfológicas entre clases (Figura 8.2).

8.4.1. Experimentos previos al clustering

Selección de la dimensión y tipo de las características

En la etapa de extracción de características realizada mediante la aproximación poligonal del ECG se obtienen una serie de T muestras de dimensión 2 (amplitud, duración) que, unidas mediante segmentos, representan la forma de onda caracterizada (en este caso los latidos del ECG). En el experimento que sigue, se tratará de determinar con qué dimensión de la señal ECG se obtienen unos resultados de clustering óptimos: si (i) utilizando únicamente las muestras de amplitud; si (ii) sólo con la duración de los segmentos; bien (iii) utilizando una combinación de ambos como la pendiente; o (iv) creando HHMs bidimensionales que evalúen conjuntamente los objetos formados por T segmentos bidimensionales. Para ello se ha seleccionado un conjunto de 100 latidos pertenecientes a dos clases distintas (54 objetos etiquetados como lati-

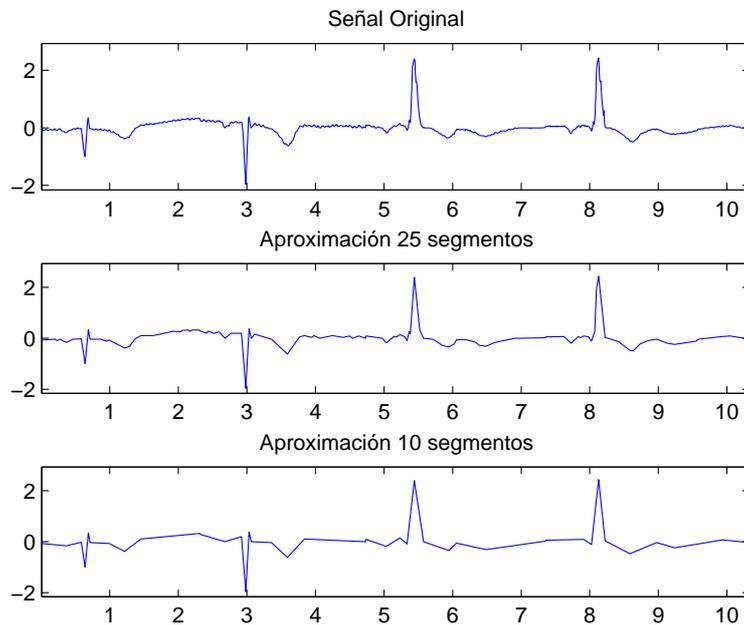


Figura 8.2: Se presentan dos morfologías de latidos caracterizadas mediante tres niveles de aproximación (la señal original, una aproximación de nivel 25 y una de nivel 10). Si comparamos visualmente entre los diferentes niveles de aproximación presentados se puede apreciar como, incluso para los objetos con un menor número de características (latidos aproximados mediante 10 segmentos) se mantienen las diferencias morfológicas entre las clases. En el eje de abscisas se representa el tiempo (s) y en el de ordenadas la amplitud (mV).

dos de tipo 1 y otros 46 objetos de tipo 2). Una vez determinados la dimensión y el tipo de las muestras a utilizar en el clustering, se reducirán las 25 características iniciales a 2 mediante la proyección de los puntos sobre un par de modelos inicializados de forma aleatoria. De esta manera, cada latido quedará caracterizado bidimensionalmente por las probabilidades de que haya sido generado por uno u otro de los dos HMMs en cuestión, con lo que la matriz de similitud entre objetos obtenida resulta de dimensión $2 \times n$ (donde n es el número de objetos). Una vez obtenida la matriz de similitud se aplicará sobre los datos un clustering con el algoritmo Max-Min, métrica por distancia

euclídea y detección automática del número de clusters. Los resultados obtenidos se presentan en la *Figura 8.3* donde, además, se ha añadido el índice de precisión del clustering obtenido en cada caso. Para una rápida estimación visual de los resultados, los latidos aparecen representados con sus etiquetas reales.

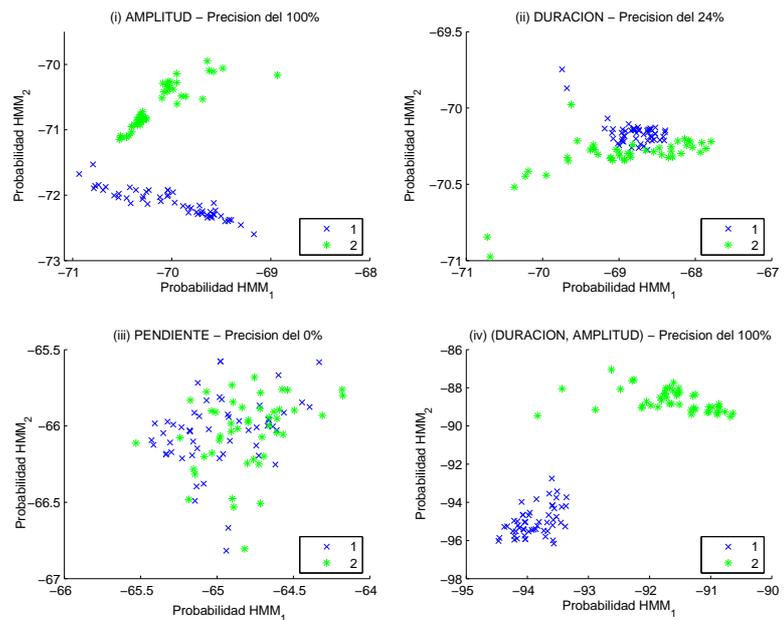


Figura 8.3: Para la selección de la dimensión y tipo de las características se han realizado cuatro pruebas de clustering utilizando los valores de (i) amplitud, (ii) duración, (iii) pendiente y (iv) el binomio (amplitud, duración). Los HMMs utilizados en la reducción de las características han sido inicializados aleatoriamente. Además se ha añadido el estimador de precisión.

Selección del tipo de modelo inicial

En este experimento se procede a la optimización del tipo de modelo a utilizar, comparando las matrices de similitud obtenidas al caracterizar los objetos utilizando combinaciones de HMMs inicializados de diferente manera.

La estimación de dos probabilidades por latido facilita en este caso su representación en un espacio bidimensional (*Figura 8.4*). La prueba se ha realizado sobre la misma señal que en el experimento anterior, utilizando en este caso la amplitud como característica unidimensional. El conjunto de modelos utilizados ha sido el siguiente y por este orden: (i) un HMM inicializado de forma aleatoria, (ii) un segundo modelo inicializado uniformemente, (iii) un tercer modelo inicializado de izquierda a derecha y finalmente (iv) un último HMM inicializado por el método de Viterbi. La interpretación de los gráficos de dispersión se entiende teniendo en cuenta el orden de los modelos citados por lo que, en la primera fila de figuras se han representado, en el eje de abscisas las probabilidades de los objetos obtenidas con el HMM aleatorio, mientras que en el eje de ordenadas aparecen las probabilidades de los latidos proyectados sobre los modelos aleatorio, uniforme, izquierda a derecha y Viterbi respectivamente. Esto significa que en las figuras ubicadas en la diagonal, el valor de probabilidad de las abscisas coincide con el de ordenadas, razón por la que los objetos se agrupan exactamente sobre la diagonal. Por cuestiones de claridad, los valores de probabilidad proporcionados por los diferentes modelos no han sido incluidos en la figura.

Selección de los centroides iniciales

Basándonos en los resultados del experimento anterior utilizaremos un modelo del tipo izquierda a derecha para, aplicando el algoritmo Max-Min, seleccionar en primera instancia al menos dos centroides de entre todos los objetos del conjunto de trabajo. Dada la facilidad para discriminar latidos de morfología diferente que presentan los modelos inicializados de izquierda a derecha (*Figura 8.4* donde, en el tercer diagrama de la diagonal en el que se reducen características con los modelos mencionados, los latidos extremos corresponden claramente a clases diferentes) comprobaremos a continuación si los objetos seleccionados como centroides en una primera iteración del clustering pertenecen o no a clases diferentes. En este caso y para validar el uso de estos modelos se procede a la representación tridimensional de una señal ECG con tres tipos de latidos diferentes. A partir de un latido de cada una de las clases se inicializan tres HMMs de izquierda a derecha sobre los que calcular las probabilidades del resto de objetos. En este caso los latidos del conjunto

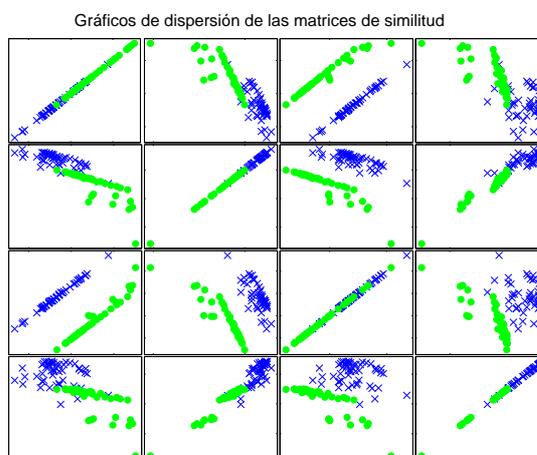


Figura 8.4: Gráficos de dispersión de las matrices de probabilidad. Para su obtención se han utilizado cuatro HMMs inicializados de distinta forma. Los valores de probabilidad logarítmica correspondientes a los latidos de tipo 1 se representan con un aspa mientras que la probabilidad logarítmica de los de tipo 2 aparecen con un punto.

seleccionado se agrupan en tres clases (44 objetos etiquetados de tipo 1, otros 38 objetos de tipo 2 y 18 latidos de tipo 3). Siguiendo las recomendaciones citadas en [Kos96], los modelos utilizados son de 25 estados (con un total de 25 segmentos por latido en el proceso de aproximación poligonal se consigue recoger la información más relevante del objeto aproximado). Los resultados se presentan en la *Figura 8.5*. En los ejes aparecen los valores de probabilidad resultantes de la proyección de los objetos sobre cada uno de los modelos. Las etiquetas utilizadas son las reales de cada latido.

Optimización del número de estados (i)

A continuación se pretenden simplificar los modelos de 25 estados utilizados en pruebas anteriores. Para ello se aplicarán dos técnicas distintas para la optimización del número de estados: (i) la propuesta en [BMF03] y que parte del número máximo de estados para, posteriormente, ir decrementando el mo-

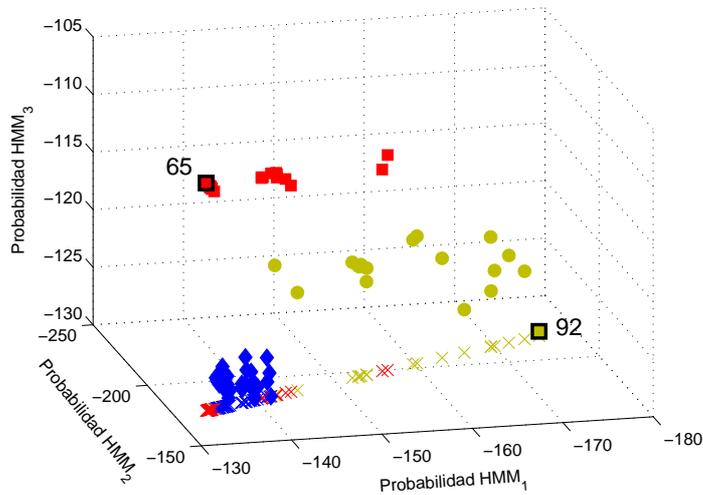


Figura 8.5: Selección de los centroides iniciales mediante tres HMMs de 25 estados inicializados de izquierda a derecha. Las etiquetas utilizadas corresponden con la clase real del latido.

delo eliminando siempre el estado que presente una menor probabilidad estacionaria (estrategia de poda); y (ii) la utilizada en [LB00] donde se eliminan estados de forma aleatoria. Para la secuencia de entrenamiento se han utilizado un total de 50 patrones bidimensionales aproximados con 15 segmentos cada uno y pertenecientes todos ellos a latidos de una misma clase. El HMM inicial se ha estimado aleatoriamente. Además de los dos algoritmos de optimización citados, para la estimación del número de estados se han aplicado dos funciones de coste distintas: el criterio MMDL (*Mixture Minimum Description Length*) y el BIC (*Bayesian Inference Criterion*). Los resultados obtenidos son los que se presentan en las Figuras 8.6 y 8.7.

Optimización del número de estados (ii)

En el experimento anterior se ha tratado de optimizar la estructura del modelo en base a las series de datos de que se dispone para el entrenamiento. Los criterios empleados han sido el BIC y MMDL, aunque los resultados

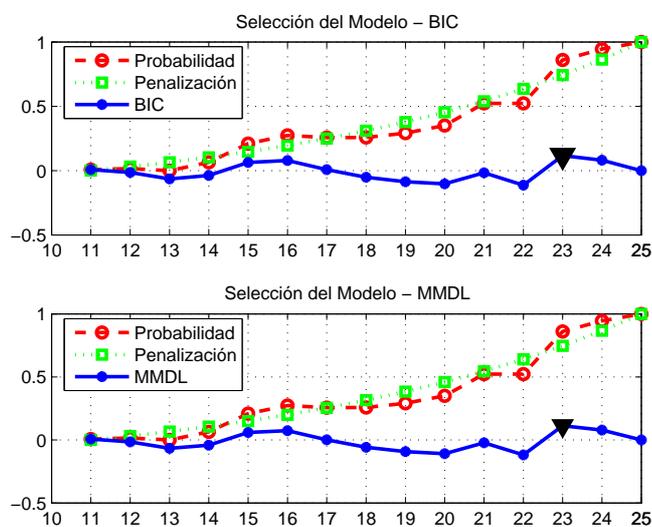


Figura 8.6: Optimización del número de estados por estrategia de poda. Se presentan los resultados obtenidos mediante la aplicación de los criterios MMDL y BIC. En el eje de abscisas aparece el número de estados del modelo evaluado, en las ordenadas el valor de la probabilidad.

obtenidos con estos no se ajustan a las apreciaciones prácticas del autor. Esto puede ser debido a que los criterios BIC y MMDL propuestos sí que optimizan en cierto sentido la estructura del modelo pero no la optimizan de cara al clustering posterior (es decir, qué modelos con estructuras *no óptimas* según el criterio proporcionen mejores resultados de sensibilidad y especificidad sobre los datos ya clasificados). Esto significa que el criterio práctico a aplicar en la optimización del número de estados debe ser establecido en base a los resultados de clustering obtenidos, y no únicamente a la vista de los datos antes de su procesamiento (tal y como se hace con la aplicación del BIC o MMDL). En este caso, y una vez se han obtenido los resultados experimentales de la *Tabla 8.1* el autor propone como óptimo aquel modelo que consiga maximizar la distancia *intercluster* minimizando al mismo tiempo la distancia *intracluster*, es decir, será mejor el modelo que consiga discriminar mejor los objetos en grupos claramente identificables. El experimento realizado consiste en modelar una morfología en base a secuencias de latidos que pertenecen a esa clase

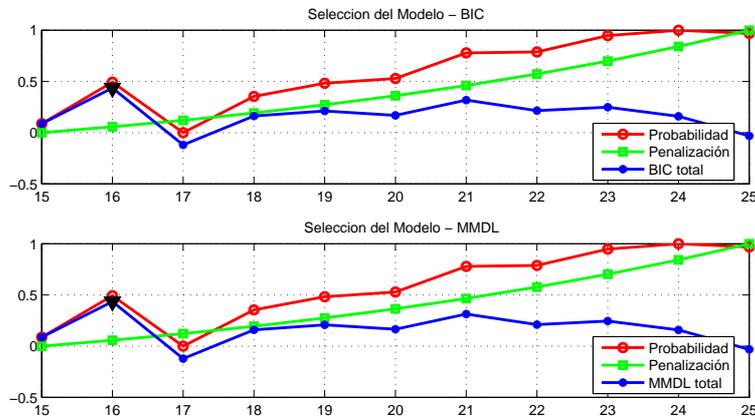


Figura 8.7: Optimización del número de estados por selección aleatoria del estado a eliminar. Se presentan los resultados obtenidos mediante la aplicación de los criterios MMDL y BIC. En el eje de abscisas aparece el número de estados del modelo evaluado, en las ordenadas el valor de la probabilidad.

para, a posteriori, utilizar el modelo en el clustering de una ECG formada por latidos de la clase modelada mezclados con otros latidos de una clase distinta (dos clusters). Los parámetros variados son tanto el número de segmentos a utilizar en la aproximación del latido como el número de estados del modelo. El poder discriminador del modelo se mide en términos de distancia intercluster, calculando la diferencia entre la probabilidad generada por el peor objeto de la clase y la del objeto clasificado fuera de la clase con máxima probabilidad. El número de estados óptimo será el presentado por el modelo con mayor distancia intercluster. Los resultados obtenidos son los del *Cuadro 8.1*.

Experimento	Segmentos aprox. pol.	Estados HMM	Distancia intercluster
1	20	30	-6,14
2	20	25	-11,78
3	15	22	151,38
4	15	20	168,25
5	15	18	80,60

Cuadro 8.1: Optimización del número de estados de un modelo en base a la distancia intercluster. Una distancia negativa indica que hay solapamiento entre los dos clusters.

8.4.2. Experimentos de clustering

Clustering con optimización de los modelos

A continuación se presenta el experimento donde se comprueba el funcionamiento de todo el proceso de clustering propuesto por el autor en la *Figura 8.1*. En este caso la señal utilizada presentaba una serie corta de 100 latidos reales aproximados con 25 segmentos y previamente segmentados, pertenecientes a dos clases distintas (54 objetos etiquetados como latidos de tipo 1 y otros 46 objetos de tipo 2). La selección del latido inicial se ha hecho de manera aleatoria y el modelado del primer HMM ha sido del tipo izquierda a derecha. Los modelos utilizados son bidimensionales, calculando la probabilidad conjunta resultante de la muestra (duración, amplitud) presentada en cada instante. Como criterio para determinar el final del clustering se ha utilizado el de precisión de los resultados fijando un valor de calidad por encima del 97%. Además, en la etapa de entrenamiento de los modelos se ha implementado el algoritmo de poda para optimización del número de estados, de tal forma que los modelos calculados siempre fueran los óptimos (aún a costa de penalizar la velocidad del proceso global debido a los sucesivos entrenamientos recibidos en una misma iteración). En cuanto a las secuencias de entrenamiento utilizadas en cada caso, han sido calculadas a partir de los 10 latidos de mayor probabilidad para cada cluster. El algoritmo de clustering utilizado ha sido el Max-Min. Para la detección del número de clusters se ha programado el método intrínseco (ver *Capítulo 6*). La evolución de las agrupaciones de objetos se puede observar en la *Figura 8.8*. Los resultados del clustering en la última de las iteraciones aparece en la *Figura 8.9*.

Clustering sobre un ECG de larga duración

En este experimento se va a desarrollar todo el proceso de clustering sobre una señal ECG real formada por un total de 3000 latidos (unos 50 minutos de duración) en la que se presentan 4 morfologías distintas. Los modelos utilizados tienen un tamaño fijo de 20 estados. Para las señales de entrenamiento se han utilizado más de 300 objetos de cada clase. La representación de los objetos (una vez reducidos a 4 características por proyección sobre los 4 modelos

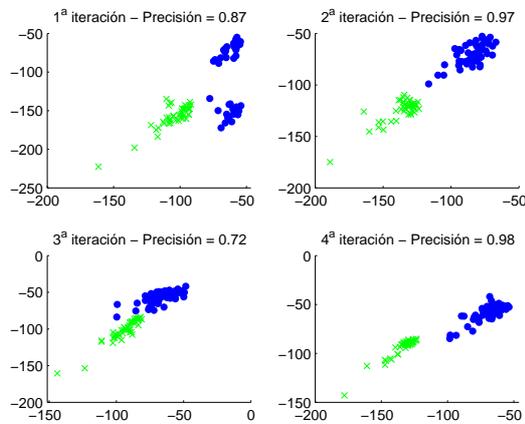


Figura 8.8: Evolución de las agrupaciones de objetos en el proceso de clustering con optimización del número de estados por poda. Las características de los latidos se han reducido a 2 mediante aplicación de HMM. Tanto en el eje de abscisas como en el de ordenadas se presentan valores de probabilidad logarítmica. Las etiquetas utilizadas (puntos para los latidos de una clase y aspas para los latidos de la otra clase) son las reales de cada latido. La precisión del clustering en cada iteración es el mostrado sobre cada una de las sub-figuras.

entrenados) es la que se muestra en la *Figura 8.10*. Las etiquetas utilizadas son las de los latidos reales. Por otra parte, en un experimento posterior se va a utilizar la matriz de similitud obtenida (de dimensión 3000×4) para comparar los resultados de clustering utilizando dos algoritmos diferentes (el Max-Min con detección automática del número de clusters y el K-medias particularizado para la creación de 4 clusters). Los resultados obtenidos se muestran en la *Figura 8.11*. Finalmente también se ha aprovechado el experimento para establecer el porcentaje de mejora obtenido con los modelos entrenados frente a los resultados del clustering utilizando modelos sin entrenar (ver *Cuadro 8.2*).

Clustering sobre una colección de latidos de gran variabilidad

En este experimento se va a aplicar el proceso de clustering descrito anteriormente sobre una colección de fragmentos de señales ECG reales obtenidas

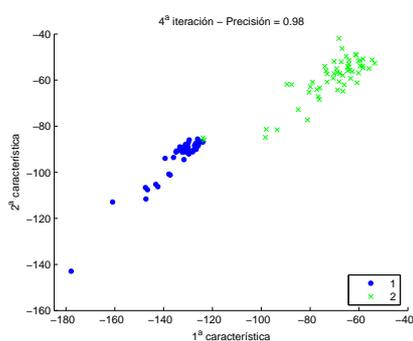


Figura 8.9: Representación de los objetos al final del proceso de clustering. Al igual que en la Figura 8.8, los valores de los ejes son probabilidades logarítmicas. Las etiquetas utilizadas en este caso son las derivadas del clustering. Sólo dos latidos son clasificados erróneamente.

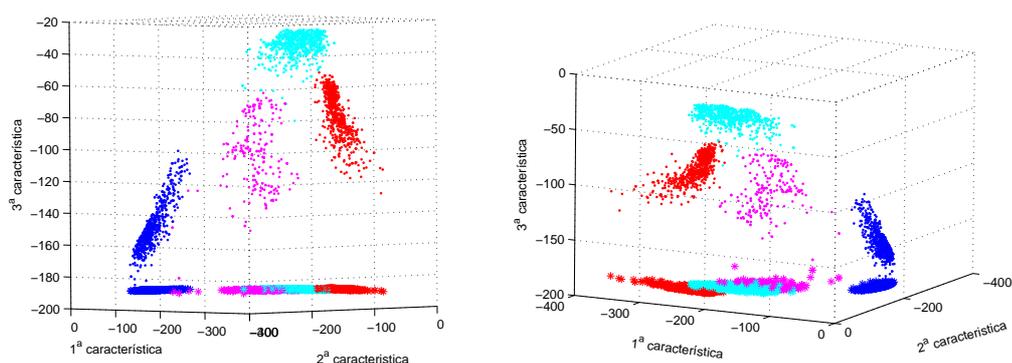


Figura 8.10: Reducción de las características de los latidos de una señal ECG real donde, a pesar de aparecer representadas sólo tres de las cuatro características extraídas, y gracias a las dos perspectivas mostradas en la figura, se aprecian claramente los cuatro grupos formados para cada una de las morfologías incluidas en el Holter. En los ejes aparece la probabilidad logarítmica.

de la base de datos del MIT [GAG⁺00]. Los resultados obtenidos se presentan en el Cuadro 8.3. Los parámetros utilizados en el proceso son los mismos que en pruebas anteriores. Cabe destacar cómo, del detalle de los clusters al final

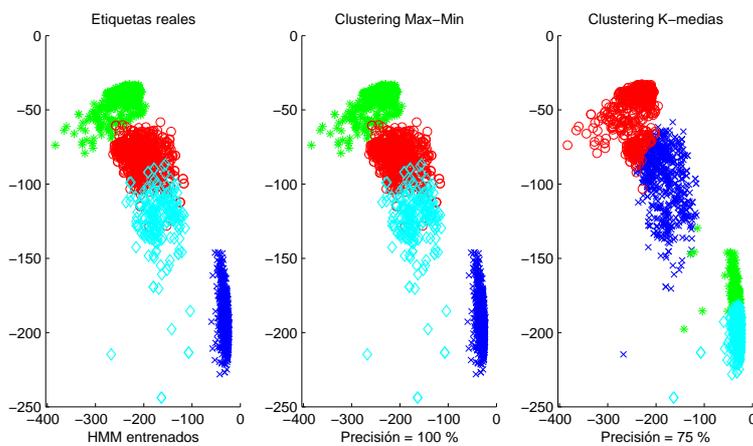


Figura 8.11: Clustering de latidos sobre los datos caracterizados en la *Figura 8.10* donde se establece la comparativa entre el algoritmo Max-Min y el K-medias. En los ejes se muestra la probabilidad logarítmica. La sub-figura de la izquierda aparecen etiquetados las morfologías reales de los latidos a agrupar.

del proceso, se observa que en la mayor parte de los casos el porcentaje de TP y TN es muy elevado y sólo empeoran el clustering los outliers que suponen un aumento en el porcentaje de FP en los clusters mayoritarios.

	Max-Min	K-medias
HMM inicial	93	81
HMM entrenado	100	75
Mejora	7	-6

Cuadro 8.2: Porcentaje de mejora obtenido con el entrenamiento de los modelos para cada uno de los algoritmos utilizados. El porcentaje de mejora se calcula como la diferencia entre los porcentajes de precisión del clustering obtenidos mediante los modelos entrenados y sin entrenar.

Fuente	Etiquetas reales	Outliers	Etiquetas de clustering	Precisión total	Iteraciones
1101	936 - 21 - 3	3	939 - 21	0,9939	2
11	251 - 64 - 40	40	252 - 74 - 29	0,8899	3
12	951 - 43 - 23	23	936 - 81	0,9682	2
100	934 - 7 - 9 - 6	6	932 - 24	0,9913	2

Cuadro 8.3: Clustering sobre una colección de latidos de gran variabilidad. En el cuadro se detalla la composición en latidos del ECG (*Etiquetas reales*) así como los grupos de latidos detectados tras el clustering (*Etiquetas de clustering*).

8.5. Conclusiones

Las pruebas anteriores han sido realizado para optimizar el proceso general de clustering de los latidos de un Holter ECG utilizando para ello HMMs. Las conclusiones derivadas de la interpretación de los resultados obtenidos en la *Sección 8.4* son las siguientes:

1. *Selección de la dimensión y tipo de las características:* donde por simple inspección visual de las agrupaciones que aparecen en la *Figura 8.3* se observa como los mejores resultados del clustering son proporcionados cuando se caracteriza unidimensionalmente con muestras de amplitud o cuando se utilizan modelos bidimensionales que proporcionan la probabilidad conjunta sobre las muestras (amplitud, duración). En ambos casos la precisión en el clustering resulta del 100 %. Queda descartado el clustering utilizando las muestras de duración o las de pendiente. De todas formas, hay que tener en cuenta que los valores de probabilidad utilizados en las matrices de similitud dependen de las características de los HMMs utilizados. En nuestro caso, para las muestras unidimensionales se utilizan los mismos modelos y para las muestras bidimensionales se recalculan los valores de las matrices de medias y covarianzas (probabilidades de emisión del modelo) para la nueva dimensión. El mayor poder discriminante de los modelos bidimensionales (se aprecian los dos grupos de latidos con menor varianza *intracluster* y mayor varianza *intercluster*) se explica por la utilización en este caso de dos características (amplitud y duración) frente a la única característica de amplitud usada en los modelos unidimensionales. Para la optimización de los resultados del clustering y a la vista de la *Figura 8.3* se aconseja que, para el

modelado de latidos, se utilicen HMMs de dos dimensiones. Finalmente añadir que el número de clusters detectado automáticamente coincide con el número real de tipos de latidos incluidos en la señal de test.

2. *Selección del tipo de modelo inicial:* por inspección visual sobre los gráficos de dispersión representados en la *Figura 8.4* se observa como, independientemente de los valores de probabilidad proporcionados, la combinación de HMMs que mejores resultados proporciona es la de un modelo inicializado de izquierda a derecha conjuntamente bien con un modelo inicializado aleatoriamente, bien con un modelo uniforme.
3. *Selección de los centroides iniciales.* De los resultados obtenidos para la selección de los centroides iniciales utilizando modelos inicializados de izquierda a derecha en la *Figura 8.5* se puede comprobar como la aplicación del algoritmo Max-Min para la selección de los centroides iniciales funciona correctamente al escoger latidos que pertenecientes a clases diferentes. Además, de un total de 25 pruebas realizadas sobre señales ECG reales, en el 100 % de los casos los centroides iniciales obtenidos han pertenecido a clases diferentes, lo que asegura la conveniencia de este tipo de inicialización de los modelos en la primera etapa.
4. *Optimización del número de estados (i) y (ii):* donde en las pruebas realizadas se han aplicado varios algoritmos y distintos criterios de selección. De los resultados obtenidos se deduce que, independientemente del criterio utilizado (BIC o MMDL seleccionan el mismo número de estados para el modelo), el número óptimo de estados del HMM entrenado con secuencias de latidos aproximadas con 15 segmentos es de 23. Frente a la variabilidad de los resultados obtenidos en la aplicación del método de [LB00] (probablemente debida a la aleatoriedad en la selección del estado a eliminar) cabe resaltar la uniformidad de resultados obtenidos con la estrategia de poda [BMF03] (donde siempre se elimina del estado de menor probabilidad estacionaria). Como el número de estados óptimo para un modelo depende tanto del número de características de las secuencias de entrenamiento como de la morfología de los objetos pertenecientes a dicha secuencia (aunque esto último está íntimamente relacionado con el número de características) no se puede, por tanto, generalizar en la utilización de un número de estados determinado. Además,

tanto en [LB00] como en [BMF03], el estado óptimo se determina cuando se llega a un compromiso entre la probabilidad ofrecida por el modelo entrenado y la penalización (que aumenta con el número de estados del modelo), es decir, cuanto más complejo es el modelo utilizado (obviamente) mejor probabilidad proporciona pero menor capacidad de generalización ofrece sobre cambios en la secuencia de datos (por lo que es penalizado). En nuestro caso consideraremos un HMM como modelo óptimo cuando mejores resultados de clustering se obtengan con él. Por este motivo y por los resultados obtenidos con los numerosos experimentos realizados (*Cuadro 8.1*) se puede concluir con que ninguno de los dos métodos de optimización de la estructura del HMM propuestos sería de aplicación válida para el caso del clustering de latidos de señales ECG. En este caso la experiencia demuestra como los mejores resultados de clustering se obtienen al utilizar secuencias de entrenamiento formadas por latidos aproximados poligonalmente con 15 segmentos para modelarlos posteriormente con HMMs de 20 estados. El utilizar modelos con un mayor número de estados incide negativamente en su generalidad (modelos demasiado específicos que no son capaces de clasificar latidos de igual morfología pero con alguna característica dispar a los de la secuencia de entrenamiento), con lo que la matriz de similitud de la clase se dispersa y la posterior aplicación del algoritmo de clustering proporciona peores tasas de sensibilidad y especificidad. El utilizar modelos con menor número de estados supone disminuir la capacidad para distinguir morfologías de su clase resultando en un aumento de la tasa de FP y FN en los clusters detectados con lo que la calidad global del proceso también disminuye.

5. *Clustering con optimización de los modelos:* tras la ejecución del experimento y tal y como se observa en la *Figura 8.8*, la calidad del proceso global se ha ido incrementando con el paso de las sucesivas iteraciones hasta llegar a obtener, en la última iteración, una precisión en el clustering del 98% habiendo identificado la totalidad de objetos de uno de los clusters y con sólo dos falsos positivos (FP) sobre ese mismo cluster. Visualmente esto se puede apreciar en la cada vez mayor distancia existente entre los dos grupos de puntos que aparecen en las gráficas que, al mismo tiempo, parecen concentrarse en torno a su centroide. El número

óptimo de estados obtenidos calculados para aproximar objetos de 25 segmentos se encuentra en torno a los 26, cifra ligeramente superior al de líneas poligonales de la aproximación, pero no excesivamente superior, en contra de lo propuesto por Koski en [Kos96] (donde se aconsejan modelos de 25 estados para latidos de 15 segmentos). Por nuestra parte y dados los resultados obtenidos en el experimento anterior, abogamos más por la línea de Koski simplificando todavía más los modelos utilizados (latidos de 15 segmentos y HMMs de 20 estados) por lo que resulta aconsejable el prescindir de la fase de optimización de los modelos, al menos en su aplicación a señales ECG. Por otro lado, la utilización de sólo 10 secuencias de datos en el entrenamiento de los modelos puede haber supuesto un inconveniente en cuanto a su capacidad de generalización final. No obstante, esto es subsanable al ejecutar el proceso global sobre series con un mayor número de datos.

6. *Clustering sobre un ECG de larga duración:* de los resultados presentados en la *Figura 8.11* podemos decir que, mediante un adecuado dimensionamiento de la secuencia de entrenamiento se pueden optimizar los modelos para cada morfología con lo que la selección de las características se reduce a la obtención de la probabilidad del objeto sobre cada uno de los modelos que componen la prueba. Tal y como se aprecia, ya en una representación tridimensional de tres de las cuatro características a las que se han reducido los objetos, éstos se presentan espacialmente discriminados con lo que se facilita la detección de los clusters. Si comparamos los resultados proporcionados por los algoritmos de clustering aplicados sobre la matriz de similitud obtenida del proceso de extracción de características podemos observar como, gracias a la excelente caracterización obtenida con los HMMs, el algoritmo que mejores prestaciones proporciona es el Max-Min (que elige como centroide el modelo de mayor probabilidad) con un estimador de precisión total del 100 %. El algoritmo K-medias en cambio, y debido a una poco acertada inicialización de los centroides (elección aleatoria de 4 centroides iniciales), no proporciona los resultados apetecidos. Es de suponer que, por su tendencia a minimizar la varianza intracluster, si utilizamos como centroides iniciales latidos de diferentes morfologías su precisión también sea próxima al 100 % [Cue01]. En cuanto al porcentaje de mejora obtenido median-

te el entrenamiento de los modelos se puede comprobar en el *Cuadro 8.2* como, aunque la calidad del proceso resulta bastante aceptable utilizando directamente los modelos iniciales sin entrenar, para el Max-Min el entrenamiento de los modelos mejora los resultados en torno al 10%. La mejora en algoritmos que, como el K-medias, no ofrecen buenos resultados, resulta inapreciable en este caso por lo que se puede concluir que el entrenamiento de los modelos no es garantía de buenos resultados si el algoritmo de clustering utilizado no es el idóneo.

7. *Clustering sobre una colección de latidos de gran variabilidad:* de los resultados presentados en el *Cuadro 8.3* y de la evolución presentada en la *Figura 8.12* se deducen las siguientes conclusiones:

- Tras la inspección visual de los fragmentos de ECG, se observa como la baja precisión del clustering presentada en alguno de los experimentos es debida principalmente a errores en la etapa de segmentación.
- En cuanto a las señales que presentan outliers, éstos se suelen localizar lejos de los clusters principales e influyen negativamente en la correcta detección del número de clusters. En la mayor parte de los casos, el porcentaje de TP y TN es muy elevado empeorando el resultado final a causa de los outliers, que suponen un aumento en el porcentaje de FP en los clusters mayoritarios. Los resultados obtenidos para el procesamiento de una misma señal sometida a una fase de eliminación de outliers mejoran perceptiblemente.
- Los resultados del clustering mejoran tras el entrenamiento de los modelos (ver *Figura 8.12*). En este caso los índices de calidad se calculan sin tener en cuenta los outliers.
- En cuanto a los algoritmos de clustering empleados, cabe destacar los buenos resultados proporcionados por el Max-Min que, aunque sensible a los outliers, lo es mucho menos que el K-medias (que, si no está correctamente inicializado recalcula centroides y tiende a disgregar los clusters).
- En cualquier caso, sería conveniente el implementar una etapa de consolidación de los clusters que correspondan realmente a una misma morfología.

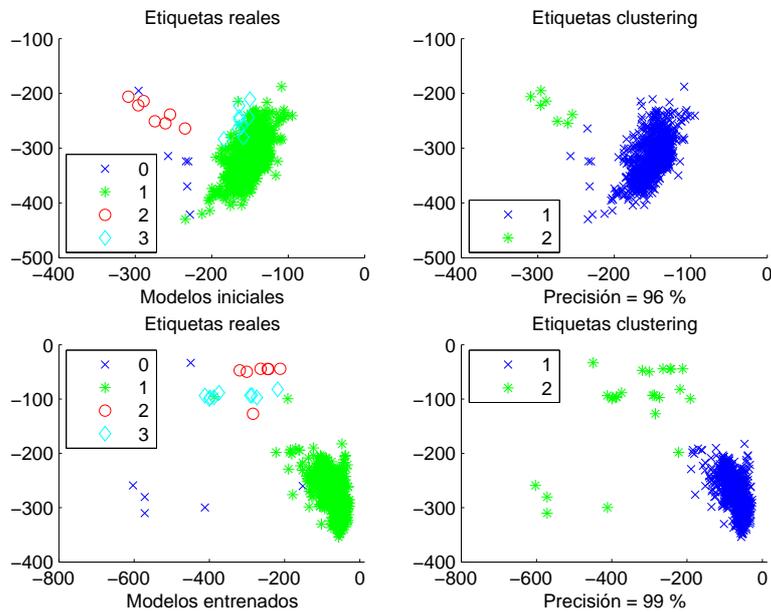


Figura 8.12: Clustering de latidos: modelos iniciales vs entrenados. En la fila superior se aprecian las etiquetas reales y de clustering para los latidos agrupados a partir de la matriz de similitud obtenida mediante los modelos inicializados de izquierda a derecha. En la fila inferior tenemos la misma representación de las etiquetas para los modelos entrenados. Tal y como se aprecia y aún siendo la precisión del clustering inicialmente buena, tras el entrenamiento de los modelos ésta aumenta en un 3%. Los índices de calidad se calculan sin tener en cuenta los outliers (etiquetados con 0).

Resumiendo los resultados obtenidos en este capítulo en una conclusión general y suponiendo que la señal ha sido preprocesada adecuadamente en etapas posteriores (la calidad de los resultados obtenidos se ve deteriorada en muchos de los casos por fallos en el etiquetado o en la etapa de segmentación) se podría decir que la reducción de las características de los latidos utilizando los *modelos ocultos de Markov* resulta muy satisfactoria ya que de su aplicación se obtiene una matriz de similitud entre los objetos donde, inicialmente, ya aparecen separados en grupos por morfologías. La calidad final del clustering va a depender, por tanto, del algoritmo utilizado, aunque los resultados presentados nos llevan a aconsejar la clasificación del latido directamente so-

bre el modelo que proporcione la máxima probabilidad (algoritmo Max-Min). Como contrapartida cabría destacar la dificultad que suponen tanto la óptima inicialización de los modelos como la adecuada obtención de secuencias de entrenamiento. Para las señales de tipo ECG (y para las señales biomédicas en general), una manera de independizar los resultados obtenidos del modelo inicial utilizado en cada caso consiste en entrenar los HMMs a partir de modelos de izquierda a derecha. Finalmente cabe resaltar que, para el caso concreto de señales biomédicas de tipo electrocardiográfico, los mayores beneficios en el clustering se obtienen al aproximar poligonalmente los latidos con secuencias de 15 segmentos para reducir posteriormente estas características mediante la utilización de modelos con 20 estados.

Parte IV

Conclusiones

Capítulo 9

Conclusiones y futuras líneas de investigación

Resumen. En este capítulo se presentan las conclusiones finales obtenidas de los experimentos realizados para el clustering de los latidos que componen una señal ECG de tipo Holter. Asimismo también se enumeran las múltiples líneas de investigación que, siguiendo con el trabajo de procesamiento y análisis automático de señales biomédicas, quedan propuestas para el desarrollo de trabajos futuros encaminados en esa misma dirección.

Contenidos

9.1. Conclusiones	225
9.2. Futuras líneas de investigación	231

9.1. Conclusiones

A continuación se resumen las conclusiones derivadas de los resultados obtenidos en las pruebas realizadas para cada uno de los capítulos que componen esta tesis:

Capítulo 3: Estado del arte

Como conclusión cabría destacar que, dejando de considerar la tesis doctoral con la que se abrió la línea de investigación del clustering de latidos de señales ECG de larga duración [Cue01], no se ha encontrado ningún otro trabajo previo al respecto de la aplicación concreta que se pretende desarrollar a lo largo de este documento: el procesamiento automático de una señal Holter ECG mediante la reducción de las características de los latidos que la componen (aplicando distintos métodos estadísticos como PCA y HMM) para su posterior clustering.

Capítulo 4: Extracción de características

En el proceso de aproximación poligonal y si comparamos los resultados de las diferentes pruebas realizadas, resulta patente la superioridad de prestaciones que, sobre el resto de métricas, presentan estimadores simples como el del error de amplitud o el error absoluto. Si a todo esto añadimos que la fidelidad de todas las aproximaciones es elevada (manteniendo hasta el 99 % de la información crítica del ECG), podemos concluir con que es el estimador por error absoluto (que mejora levemente las prestaciones ofrecidas por el error de amplitud, en cuanto a tiempo de procesamiento y CRP se refiere) el que optimiza el proceso de cálculo de la aproximación poligonal de señales ECG de larga duración, permitiendo las aproximaciones sin pérdida de información crítica y obteniendo ratios de compresión del 50 % si el umbral de error es mínimo y de hasta el 76 % con un umbral de error relajado en un 500 %. Además, la justificación de una buena aproximación poligonal se demuestra actuando en el procesamiento de tareas complejas donde se comprueba el incremento del ratio entre la velocidad de procesamiento cuando se actúa sobre

la señal comprimida y la velocidad de procesamiento cuando se actúa sobre la señal original.

Capítulo 5: Preprocesamiento y segmentación de la señal ECG

Tras las múltiples actuaciones que sobre el ECG se realizan en este capítulo podemos concluir en lo siguiente: (i) para la eliminación de las interferencias introducidas por la red eléctrica se procede inicialmente a un filtrado selectivo de la señal; (ii) las variaciones de la línea de base se eliminan mediante wavelets; (iii) para la eliminación del ruido interferente se aprovecha el propio proceso de aproximación poligonal; (iv) la segmentación del ECG en latidos también se realiza sobre la aproximación poligonal obteniendo hasta un 65 % de ahorro en el tiempo de procesamiento.

Capítulo 6: Introducción al clustering

Como conclusión a este capítulo teórico cabría resaltar la importancia que, en el proceso global, suponen tanto la adecuada extracción de las características de los elementos a agrupar como la cuidada selección de la función de coste o criterio de clustering. Además, también se deberá tener en consideración el problema añadido de los outliers, cuyo adecuado tratamiento puede suponer una mejora importante de los resultados finales del clustering.

Capítulo 7: Clustering mediante Análisis de Componentes Principales

Tras efectuar los distintos experimentos de clustering utilizando PCA en la etapa de reducción de características y en cuanto al comportamiento de los clusters se refiere se puede interpretar lo siguiente: el hecho de la aparición de grandes clusters frente a clusters muy pequeños (y cuando el número total de clusters es reducido) combinado con una inadecuada selección de los centroides iniciales explicaría el mal comportamiento de un algoritmo como el K-medias (que trata siempre de *forzar* nuevas agrupaciones buscando precisamente sobre el cluster mayoritario) con lo que la precisión de los clusters identificados mediante este algoritmo al final del proceso no es la esperada. En

este caso y siempre que los clusters minoritarios aparezcan lo suficientemente alejados del mayoritario, un algoritmo particional como el Max-Min establecerá como centroides de los nuevos clusters a los elementos más distantes del centroide original, reagrupando los objetos a posteriori y presentando unos resultados adecuados. Esta estrategia permite identificar los clusters pequeños con morfologías poco similares a la forma de onda de los latidos mayoritarios y hace del Max-Min el algoritmo ideal para aplicar en la detección de outliers. Lo mismo sucede con el clustering de tipo jerárquico binario, donde la inconsistencia entre pares de latidos nos lleva a agrupar los latidos morfológicamente similares, identificando los clusters minoritarios con un coeficiente de inconsistencia elevado respecto al cluster mayoritario. Es decir, el algoritmo K-medias funciona bien si la señal analizada no presenta outliers mientras que el Max-Min o el binario sí que detectan mejor los latidos de clusters minoritarios (outliers). Finalmente cabría concluir que, independientemente del algoritmo de clustering utilizado y para optimizar los resultados del clustering, la matriz de datos configurada por los latidos que mejores resultados proporciona debe estar formada por la característica de pendiente (frente a las matrices formadas sólo por la amplitud o sólo por la duración), mientras que la métrica aplicada debería ser la distancia *Euclídea* con lo que, y según la dificultad del problema planteado (número de clusters reales que presente la estructura de datos y aparición o no de outliers), la precisión en tareas de clustering simples puede llegar al 90 %.

Capítulo 8: Clustering utilizando Modelos Ocultos de Markov

Los mejores resultados de selección de las características de los latidos se obtienen aplicando los HMM sobre los puntos bidimensionales (duración, amplitud) de la aproximación poligonal. Para la inicialización de los modelos y selección de los centroides en una primera iteración se recomienda utilizar HMMs del tipo izquierda a derecha. El número de estados óptimo va a depender tanto del número de características de los objetos como de la morfología modelada pero la experiencia demuestra que los mejores resultados en el clustering posterior (y para el caso concreto de señales electrocardiográficas) se obtienen con modelos de 20 estados aplicados sobre latidos aproximados poligonalmente con 15 segmentos por lo que, en este caso en concreto, se

recomienda prescindir de la fase de optimización de los modelos propuesta en [LB00]. Una vez analizados los modelos y como conclusión general a los resultados de clustering obtenidos diremos que la reducción de las características de los latidos aplicando *modelos ocultos de Markov* resulta muy satisfactoria ya que proporciona una matriz de similitud entre los objetos donde, ya de por sí y antes de aplicar algún algoritmo de clustering, aparecen claramente separados en grupos por morfologías. Los resultados finales del clustering van a depender, por tanto, de las características del algoritmo seleccionado, aunque se aconseja agrupar el latido en el cluster representado por el HMM con máxima probabilidad. Además, al reducir las características de los objetos (reducción del número de parámetros libres) se simplifica la tarea de clasificación. Como contrapartida cabría destacar tanto la dificultad que supone la óptima inicialización de los modelos como la importancia de una correcta preparación de las secuencias de datos a utilizar en el entrenamiento de los HMMs. Dimensionando convenientemente ambos parámetros se obtienen unas precisiones en la tarea de clustering de hasta el 100 %. También tenemos que mencionar la negativa incidencia que los outliers tienen sobre la tarea de clustering, que depende del algoritmo seleccionado y no del método de reducción de características (aunque mediante HMMs los outliers son aislados de forma más precisa que utilizando otros métodos).

Conclusiones generales

De todo lo dicho anteriormente vamos a extraer una serie de conclusiones generales a modo de resumen y que son las que se exponen a continuación. Por una parte, podemos *culpar* a la especial morfología de los agrupamientos presentados por las señales ECG como la responsable de los resultados obtenidos. Efectivamente, en una señal ECG de tipo Holter aparecen un elevado número de latidos considerados como *normales* que presentan una misma morfología (hasta un 95 % del total) y donde son una minoría los latidos que presentan morfologías diferentes (bien debido a un preprocesado deficiente en el que no se han podido eliminar todas las señales interferentes, bien porque se trata de latidos *anormales* cuya forma de onda distinta se debe a motivos fisiológicos). Para comprobar esto y sea cual sea la caracterización de los latidos, se puede realizar un simple test visual representando tridimensio-

nalmente tres de las características principales de los objetos y comprobando como, normalmente, aparece un cluster mayoritario (de estructura elipsoidal) donde se aglutinan los latidos con morfología normal, y un número de clusters formados por una pequeña cantidad de latidos (en muchos de los casos se trata de un único latido, corrupto, denominado outlier) donde se presentan el resto de morfologías. La justificación del proceso de clustering aparece clara si tenemos en cuenta que, para el diagnóstico de una enfermedad cardíaca, no resulta de especial importancia el análisis de los latidos del cluster mayoritario (que son los normales) centrandolo el interés médico en aquellos pocos latidos (o series de latidos) que no han sido agrupados con la mayoría (que suelen ser los latidos patológicos), independientemente del número de clusters en que hayan sido agrupados. De esta manera, y si el médico o especialista trabaja sobre los resultados del clustering presentados a lo largo de la tesis (en lugar de hacerlo directamente sobre la señal Holter ECG), se le facilita el análisis de la señal para su diagnóstico final, ya que se consiguen eliminar del Holter una enorme cantidad de elementos cuyo análisis visual quedará certificado con el chequeo de un latido representativo del cluster.

Por otra parte y para el correcto diagnóstico realizado sobre los datos resulta imprescindible que, al finalizar el proceso de clustering, se hayan obtenido, al menos, tantos agrupamientos como morfologías distintas de latidos se encuentren en la señal analizada. Aprovechando una de las propiedades de los registros ECG (como la de presentar los latidos de una misma morfología agrupados de manera continua en una secuencia temporal) se consigue que, aún habiendo identificado una determinada forma de onda en varios clusters diferentes (si en el clustering no se consigue una correspondencia unívoca entre la forma de onda y el cluster asociado), la inspección por parte del especialista de un latido característico de cada cluster le llevará (en el peor de los casos) a volver a visualizar repetidas veces el mismo fragmento de señal ECG. En el extremo opuesto nos encontraríamos ante el caso de haber asociado de manera correcta en un mismo cluster determinada morfología que, en el Holter, aparece en secuencias temporales diferentes (por ejemplo, una arritmia producida al inicio del Holter y la repetición del mismo fenómeno varias horas después). Este hecho presenta el problema añadido de tener que identificar, dentro de un mismo cluster, secuencias de latidos temporalmente distanciadas hecho que, por otra parte, puede resultar de gran utilidad para

el diagnóstico posterior por parte del cardiólogo.

En cuanto al objetivo de ayuda diagnóstica del registro Holter, y con el fin de proporcionar dicho soporte, el proceso automático no debería terminar en la identificación de los agrupamientos sino que, además y en una etapa posterior, se deberían analizar los clusters tratando de identificar los latidos representativos con las cardiopatías correspondientes (utilizando, por ejemplo, una biblioteca de cardiopatías).

Si comparamos los métodos de reducción de características utilizados en este trabajo se puede concluir que, en principio, es preferible la utilización de los HMMs frente a la de PCA. ¿El motivo? Mientras que la técnica de modelado es muy resistente al ruido y a las interferencias generando características de tipo probabilístico (sólo hay que comparar y decidir) además de permitir el modelado de formas de onda de baja amplitud (onda P) y de proporcionar un algoritmo de estimación de los parámetros a partir de secuencias de entrenamiento, la aplicación de PCA simplemente supone una reducción de la información mediante una transformación geométrica del espacio de características sin aportar ningún otro valor añadido. PCA elimina redundancia pero depende en exceso de la información generada en el proceso de aproximación poligonal frente a la potencia de clasificación de los HMM que no dependen tanto de la fase de aproximación poligonal. Una aproximación poligonal *diferente* para latidos de la misma clase genera clusters distintos con PCA mientras que, esos mismos latidos, tienen mayores posibilidades de ser clasificados en el mismo cluster si utilizamos HMM.

Finalmente y tratando de resumir todo el trabajo realizado en una única conclusión, tenemos obligatoriamente que destacar la indudable utilidad médica que el clustering de latidos de señales ECG de larga duración supone para los cardiólogos, sin dejar de mencionar por otra parte la amplia serie de posibilidades de investigación descubiertas en el campo de la biomedicina por lo que se espera que, en un futuro inmediato, todas estas propuestas deriven en una mejora y en una aplicación real de los métodos de inspección tradicionales utilizados hasta la fecha en muchas disciplinas médicas.

9.2. Futuras líneas de investigación

A continuación, y fruto del trabajo desarrollado en esta tesis doctoral, enumeraremos tanto las líneas de investigación abiertas como todos aquellos trabajos susceptibles de ser continuados y que pueden cristalizar en futuros proyectos de colaboración con entidades médicas y sanitarias:

Compresión de la señal ECG en tiempo real

De los resultados proporcionados en el *Capítulo 4* se abre la posibilidad de investigar sobre el desarrollo de un dispositivo Holter que, en lugar de adquirir la señal ECG original, calcule directamente su aproximación poligonal (con el consiguiente ahorro en espacio de almacenamiento que ello supondría). Para ello se proponen dos alternativas: (i) *el desarrollo de un sistema de registro Holter con un microprocesador de altas prestaciones sobre el que implementar los algoritmos de aproximación y (ii) la aproximación poligonal de la señal con un umbral de error más relajado*. Utilizando esta última opción y aumentando el umbral de error en un 500 % se consiguen incrementos de la velocidad de procesamiento en el cálculo de la aproximación poligonal de hasta un 52 %, manteniendo la información crítica. Con estos parámetros y en las condiciones de un ritmo cardíaco normal, podemos comprimir un latido en 0.34s (el 45 % del tiempo de adquisición del latido), porcentaje que podría permitir la compresión de la señal ECG en tiempo real dotando al Holter de un buffer adaptado a las necesidades de la adquisición y con un microprocesador capaz de procesar la señal en los intervalos de 2.7 ms (a 360 Hz) que transcurren entre la adquisición de una muestra y la siguiente, mediante el mecanismo de interrupciones adecuado. Además, si utilizamos una técnica de aproximación por ráfagas (dedicando cierto porcentaje del tiempo del microprocesador a la tarea de aproximación en zonas del ECG con mucha información redundante) se podría asegurar que, aún perdiendo algunas muestras, no hay pérdida real de información crítica. Efectivamente, si observamos la morfología de un latido (*Capítulo 2*) podemos concluir que, al iniciarse la actividad cardíaca en una zona eléctricamente plana (antes de la excitación auricular responsable de la onda P) y si el algoritmo utilizado para la segmentación de los latidos procesa la mayor parte de la señal en este intervalo de tiempo, la pérdida de informa-

ción crítica resultante del proceso de aproximación poligonal en tiempo real resulta prácticamente nula. En este caso, la reducción del espacio de almacenamiento en el dispositivo de memoria secundaria puede llegar a ser hasta de un 76 % sobre el espacio utilizado inicialmente por la señal original.

Segmentación de latidos por comparación de patrones mediante HMM

Otra posibilidad que, a la vista los resultados y desarrollos implementados queda abierta, es la de utilización de los HMM en la etapa de segmentación de la señal ECG en latidos. En el *Capítulo 3* se proponen una serie de algoritmos basados en la comparación de patrones para la detección del complejo QRS. En este caso lo que se propone es el modelado del complejo mediante un HMM que permita la evaluación de tramos de la señal ECG en términos de probabilidad. La detección del complejo QRS de un latido corresponderá al tramo de mayor probabilidad con lo que, a partir del complejo, se procederá a la segmentación del ECG. Para la validación del nuevo método quedará pendiente la evaluación de sus prestaciones en términos de velocidad de procesamiento y coste computacional.

Aplicación de otros métodos estadísticos al clustering de latidos

Una futura línea de investigación, motivo de una tesis doctoral, queda abierta con la aplicación de los conocidos como métodos estadísticos multivariantes en el clustering de latidos de la señal ECG Holter. En este caso la aplicación de los nuevos métodos se efectuaría, no para una reducción de las características de los objetos, sino ya en la etapa de clustering sobre las distancias entre los objetos que conforman la matriz de similitud. Uno de los métodos propuestos es el de análisis estadístico multivariante de las varianzas (Multivariate Analysis of Variance, MANOVA) que aborda el análisis de la varianza de los diferentes grupos de datos y que consiste en la búsqueda de aquella combinación lineal de las variables originales que maximice la separación entre los grupos de datos originales [Mat01]. El proceso global se iniciaría con el preprocesamiento, segmentación y caracterización de la señal ECG, continuaría con una adecuada reducción de las características (mediante

PCA o HMM) y finalizaría con el clustering mediante la aplicación alguna de las técnicas de análisis estadístico multivariante existentes.

Clasificación de latidos en tiempo real

Continuando con el tema de aplicación de los HMMs, se propone la posibilidad del entrenamiento de HMMs off-line, supervisando y refinando la calidad de los modelos resultantes. De esta forma el problema de clustering quedaría reducido a un problema de clasificación sobre los modelos obtenidos en el laboratorio. La clasificación del latido adquirido se obtendría mediante el cálculo de las probabilidades sobre cada uno de los HMMs, identificando (etiquetando) finalmente el latido sobre modelo que proporciona una mayor probabilidad (siempre que esta probabilidad permanezca por encima de un umbral determinado). Además, se podría optimizar el proceso clasificatorio para intentar el etiquetado del objeto en tiempo real. Dependiendo de la capacidad del procesador utilizado, la compresión, segmentación, selección de características y etiquetado podría realizarse directamente en el dispositivo Holter de adquisición de la señal ECG [SCSA04b].

Biblioteca de HMMs para ayuda al diagnóstico

Por otra parte y de manera independiente a la propuesta anterior (clasificación en tiempo real), se podría pensar en la generación de una biblioteca de HMMs de las diferentes y variadas cardiopatías existentes. Como en muchos casos resulta difícil el segmentar correctamente los latidos que aparecen agrupados en series de difícil interpretación (latidos bigeminados o trigeminados, etc.), la creación de una biblioteca de HMMs cardiopáticos aportaría como valor añadido el poder detectar e interpretar correctamente (etiquetar correctamente) incluso las series de latidos agrupadas y segmentadas irregularmente reduciendo en la mayoría de los casos el problema del clustering que aparece con los outliers. Además, su utilización para ayuda al rápido diagnóstico podría derivar en una aplicación de diagnóstico automático de enfermedades cardiológicas (diagnóstico siempre refrendado por la opinión del especialista) que facilitaría mucho la labor de los médicos, sobre todo en el diagnóstico de cardiopatías poco frecuentes (la proyección del latido sobre la biblioteca de

modelos, tan amplia como se desee, generaría de forma automática una serie de probabilidades sobre los distintos items de la biblioteca que el cardiólogo debería interpretar posteriormente).

Modelado de nuevas cardiopatías

El capítulo anterior (*Capítulo 8*) finaliza con una nueva vía de investigación abierta, donde el autor (conjuntamente con un equipo de cardiólogos del hospital La Fe, en Valencia) participa en un proyecto (solicitado) que pretende la modelización de cierto tipo de morfologías patológicas para tratar de explicar el síndrome de muerte súbita del lactante (a través del estudio genético del síndrome de *QT* corto). En este proyecto y mediante los modelos obtenidos en el laboratorio a partir de secuencias de latidos que presenten la patología descrita, se intentan explicar los mecanismos fisiológicos responsables de los síndromes citados, y una de las opciones para ello pasa por utilizar HMMs entrenados convenientemente.

Desarrollo de un paquete de software para procesamiento de señales ECG

Tratando de sacar el máximo partido al conjunto de utilidades desarrollado para el tratamiento automático de señales de tipo biomédico, se pretende la implementación de un paquete de software que aglutine todas las anteriores aplicaciones y que permita al especialista la obtención de resultados de la manera más simple posible [CMA⁺03]. Para ello será necesaria la utilización de un lenguaje de programación que proporcione unas prestaciones profesionales (como C++). En cualquier caso, la decisión final sobre el diagnóstico de la enfermedad recaerá siempre en el criterio del especialista, que será el que deba validar los resultados ofrecidos por el software y nunca tomar éstos como diagnóstico final.

Bibliografía

- [AAKH⁺99] I.C. Al-Aweel, K.B. Krishnamurthy, J.M. Hausdorff, J.E. Mietus, J.R. Ives, A.S. Blum, D.L. Schomer, and A.L. Goldberger. Post-ictal heart rate oscillations in partial epilepsy. *Neurology*, 53(7):pp. 1590–1592, 1999.
- [AB95] J.C. Aningbogu and A. Belaid. Hidden markov models in text recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 9(6):925–958, 1995.
- [AEH99] K. Aas, L. Eikvil, and R. B. Huseby. Applications of hidden markov chains in image analysis. *Pattern Recognition*, 32(4):703–713, 1999.
- [AH98] L.M Arslan and J.H.L. Hansen. Selective training for hidden markov models with application to speech classification. *IEEE Transaction on Speech and Audio Processing*, 7(1):46–54, 1998.
- [Alb83] P. Albrecht. S-t segment characterization for long-term automated ecg analysis. MIT Dept. of Electrical Engineering and Computer Science, 1983.
- [AM99] T. Aluja and A. Morineau. *Aprender de los datos: Analisis de Componentes Principales*. Ediciones Universitarias Barcelona, 1999.
- [AMO⁺03] R. Almeida, J.P. Martínez, S. Olmos, A.P. Rocha, and P. Laguna. Automatic delineation of *t* and *p* waves using a wavelet-based multiscale approach. *1st International Congress on Computational Bioengineering*, pages pp. 219–224, 2003.

- [ARCG05] A.J.Kozakevicius, C. Ramos, R. Ceretta, and R. Guerra. Adaptive ecg filtering and qrs detection using orthogonal wavelet transform. *Biomedical Engineering Proc.*, 458, 2005.
- [ARD94] K. Anant, G.H. Rodrigue, and F. Dowla. Detection of the electrocardiogram p-wave using wavelet analysis. *Proceedings of SPIE International Symposium on OE / Aerospace Sensing*, pages pp.4–8, 1994.
- [AS85] J.A. Van Alsté and T.S. Schilder. Removal of base-line wander and power-line interference form the ecg by an efficient fir filter with a reduced number of taps. *IEEE Transactions on Biomedical Engineering*, 32(12):pp. 1052–1060, 1985.
- [AT85] M.L. Ahlstrom and J. Tompkins. Digital filters for real-time ecg signal processing using microprocessors. *IEEE Transactions on Biomedical Engineering*, 32(9):pp. 708–713, 1985.
- [Ati90] A.F. Atiya. An unsupervised learning technique for artificial neural networks. *Neural Networks*, 3:707–711, 1990.
- [Att00] H. Attias. *A Variational Bayesian Framework for Graphical Models*, volume 12 of *MIPS*. MIT Press, 2000.
- [Bak75] J.K. Baker. The dragon system-an overview. *IEEE Transaction of Accoustic Speech Signal Processing*, 23(1):24–29, 1975.
- [BB00] P. Baldi and S. Brunak. *Bioinformatics*. The MIT Press, 2000.
- [BC94] P. Baldi and Y. Chauvin. Smooth on-line learning algorithms for hidden markov models. *Neural Computation*, 6(2):305–316, 1994.
- [BC96] P. Baldi and Y. Chauvin. Hybrid modeling, hmm/nn architectures, and protein applications. *Neural Computation*, 8(6):1541–1565, 1996.
- [BC03] M.J. Beal and Z. Ghahramani C.E.Rasmussen. The infinite hidden markov model. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*. MIT Press, 2003.

- [Ber99] A. Berchtold. High-order extensions of the double chain markov model. Technical Report WA 98195-4322, Department of Statistics, University of Washington, August 1999.
- [Ber02] P. Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, Inc., 2002.
- [Bez96] A. Bezarienas. Selective noise filtering of high resolution ecg through wavelet transform. *Computers in Cardiology*, pages pp. 637–640, 1996.
- [BF95] Y. Bengio and P. Frasconi. An input-output hmm architecture. *Advances in Neural Information Processing Systems*, 7:427–434, 1995.
- [BH92] D.E. Brown and C.L. Huntley. A practical applications od simulated annealing to clustering. *Pattern Recognition*, 25(4):401–412, 1992.
- [BHH97] M. Bahoura, M. Hassani, and M. Hubin. Dsp implementation of wavelet transform for real time ecg waveforms detection and hearr rate analysis. *Computer Methods and Programs in Biomedicine*, 52:pp. 35–44, 1997.
- [Bil98] J.A. Bilmes. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical report, International Computer Science Institute Berkeley, 1998.
- [Bis95] C.M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [BM90] H. Boulard and N. Morgan. A continuous speech recognition system embedding mlp into hmm. *Advances in Neural Information Processing Systems*, 2:186–193, 1990.
- [BMF03] M. Bicego, V. Murino, and M. Figueiredo. A sequential pruning strategy for the selection of the number of states in hidden markov models. *Pattern Recognition Letters*, 24(9-10):1395–1407, June 2003.

- [BMR91] H. Boulard, N. Morgan, and S. Renals. Neural nets and hidden markov models: Review and generalizations. *Speech Communications*, 11:237–246, 1991.
- [BP66] L.E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state markov chain. *Annal of Mathematical Statistic*, 37:1554–1563, 1966.
- [Bra96] M. Brand. Coupled hidden markov models for modelling interactive process. Technical report, MIT Lab for Perceptual Computing, 1996.
- [Bur97] S. Burrus. *Introduction to Wavelets and Wavelet Transforms*. Prentice Hall, 1997.
- [BWF98] G. Biswas, J. Weinberg, and D. Fisher. Conceptual clustering algorithm for dataming applications. *IEEE Transactions on Man, Systems and Cybernetics*, 1998.
- [CAP99] D. Cuesta, G. Andreu, and J.C. Pérez. Detección del período en señales holter para su aplicación a la clasificación de patologías coronarias. In *II Jornadas de Procesamiento de la Señal, Gandía*, 1999.
- [CC98] W.T. Cheng and K.L. Chan. Classification of electrocardiogram with hidden markov models. In *Proceedings of 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 143–146, 1998.
- [CDA94a] E.J. Ciaccio, S.M. Dunn, and M. Akay. Biosignal pattern recognition and interpretation systems. part 3 of 4: Methods of classification. *IEEE Engineering in Medicine and Biology*, February–March:129–134, 1994.
- [CDA94b] E.J. Ciaccio, S.M. Dunn, and M. Akay. Biosignal pattern recognition and interpretation systems. part 4 of 4: Review of applications. *IEEE Engineering in Medicine and Biology*, April–May:269–283, 1994.
- [Cha94] J.K. Chang. Dynamic linear models with markov-switching. *Journal of Econometrics*, 60:pp. 1–22, 1994.

- [CM02] D. Cuesta and P. Micó. Clustering electrocardiograph signals using hidden markov models. *European Medical and Biological Engineering Conference*, 2002.
- [CMA⁺03] D. Cuesta, P. Micó, M. Aboy, D. Novák, R. Brezny, L. Samblas, D. Pastor, and S. Sancho. Biosignal laboratory: A software tool for biomedical signal processing and analysis. *25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2003.
- [CMNA04] D. Cuesta, P. Micó, D. Novák, and M. Aboy. Pattern matching techniques applied to biomedical signal processing. *IIAS-Transactions on Systems Research and Cybernetics*, IV(1), 2004.
- [CN02] D. Cuesta and D. Novák. Automatic extraction of significant beats from a holter register. In *The 16th international EURASIP conference BIOSIGNAL 2002*, pages 3–5, 2002.
- [CNE⁺00] D. Cuesta, D. Novák, V. Eck, J.C. Perez, and G. Andreu. Electrocardiogram baseline removal using wavelet approximation. In *BIOSIGNAL 2000 proceedings*, pages 136–138, 2000.
- [CNPA02] D. Cuesta, D. Novák, J. C. Perez, and G. Andreu. Feature extraction methods applied to the clustering of electrocardiographic signals: A comparative study. In *International Conference on Pattern Recognition, CPR-2002*, August 2002.
- [Cop68] E.T. Copson. Metric spaces. *Cambridge University Press*, 1968.
- [Cro98] M.S. Crouse. Wavelet-based statistical signal processing using hidden markov models. *IEEE Transaction on Signal Processing*, 46(4):886–902, 1998.
- [CS96] P. Cheeseman and J. Stutz. Bayesian classification (autoclass): Theory and results. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *In Advances in Knowledge Discovery and Data Mining*, pages 153–180. MA: MIT press, cambridge edition, 1996.

- [CSC90] D.A. Coast, R.M. Stern, and G.G. Cano. An approach to cardiac arrhythmia analysis using hidden markov models. *IEEE Transactions on Biomedical Engineering*, 37(9):826–836, 1990.
- [Cue99] D. Cuesta. Revisión de métodos y algoritmos para el tratamiento de señales electrocardiográficas. Dea, DISCA - UPV, 1999.
- [Cue01] D. Cuesta. *Estudio de métodos para procesamiento y agrupación de señales electrocardiográficas*. PhD thesis, Dept. Informática de Sistemas y Computadores - U.P. Valencia, 2001.
- [CW99] R.R. Coifman and M.V. Wickerhauser. Experiments with adapted wavelet de-noising for medical signals and images. Technical report, Department of Mathematics, Yale University,, November 1999.
- [CZO98] H.J. Choi, S.J. Zun, and Z.H. Oh. Robust estimation of discrete hidden markov model parameters using the entropy-based feature-parameter weighting and source-quantization modeling. *Artificial Intelligence in Engineering*, 12:243–252, 1998.
- [dCMR99] P. de Chazal, G. McDarby, and R.B. Reilly. A wavelet based classifier of the electrocardiogram. In *Proceedings of the European Medical and Biological Engineering Conference EMBEC99*, pages 540–541, 1999.
- [DGL96] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, 1996.
- [DH73] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.
- [DHS01] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley & Sons, 2001.
- [DJ95] D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Stat. Assoc.*, 90, 1995.
- [DKPB01] H. Dinh, D. Kumar, N. Pah, and P. Burton. Wavelets for qrs detection. *23rd IEEE/EMBS Conf. Proc.*, 2001.

- [DLR77] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum-likelihood from incomplete data via the em algorithm. *Journal of Royal Statistics Society*, 39(1):pp. 1–38, 1977.
- [Don96] D.L. Donoho. De-noising by soft-thresholding. Technical report, Dept. of Statistics, Stanford University, 1996.
- [DPH93] J.R. Deller, J.G. Proakis, and J.H.L. Hansen. *Discrete-Time Processing of Speech Signals*. McMillan Publishing Company, 1993.
- [ea97] M.Jansen et al. Generalized cross validation for wavelet thresholding. *Signal Processing*, 56(1):pp. 33–44, January 1997.
- [FB94] M. Ferdjallah and R.E. Barr. Adaptive digital notch filter design on the unit circle for the removal of powerline noise from biomedical signals. *IEEE Transactions on Biomedical Engineering*, 41(6):pp. 529–536, 1994.
- [FJJ⁺90] G.M. Friesen, T.C. Jannet, M.A. Jadallah, S.L. Yates, S.R. Quint, and H.T. Nagle. A comparison of the noise sensitivity of nine qrs detection algorithms. *IEEE Transactions on Biomedical Engineering*, 31(1):pp. 85–98, 1990.
- [FST98] S. Fine, Y. Singer, and N. Tishby. The hierarchical hidden markov model: Analysis and applications. *Machine Learning*, pages 32–41, 1998.
- [Fu82] K.S. Fu. *Syntactic Pattern Recognition and Applications*. Prentice-Hall, 1982.
- [Fuk90] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2 edition, 1990.
- [FXZ92] D. Fisher, L. Xu, and N. Zard. Ordering effects in clustering. In *Ninth International Conference on Machine Learning*, pages 163–168, 1992.
- [GAG⁺00] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.Ch. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, and H.E. Stanley. Physiobank, physiotoolkit and physionet: Components

- of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000. Circulation Electronic Pages: <http://circ.ahajournals.org/cgi/content/full/101/23/e215>.
- [Gal00] M.J. Gales. Cluster adaptive training of hidden markov models. *IEEE Transaction on Speech and Audio Processing*, 8(4), 2000.
- [GAMM85] S.D. Greenwald, P. Albrecht, G.B. Moody, and R.G. Mark. Estimating confidence limits for arrhythmia detector performance. *Computers in Cardiology*, 12:pp. 383–386, 1985.
- [GFP89] F. Gritzali, G. Frangakis, and G. Papakonstantinou. Detection of the p and t waves in an ecg. *Computers and Biomedical Research*, 22:pp. 83–91, 1989.
- [GG84] S. Geman and D. Geman. Stochastic relaxations, gibbs distributions and the bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.
- [GG92] A. Gersho and R.M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.
- [Gho02] J. Ghosh. Scalable clustering methods for data mining. *Handbook of Data Mining*, 2002.
- [GJ97] Z. Ghahramani and M.I. Jordan. Factorial hidden markov models. *Machine Learning*, 29:245–273, 1997.
- [GL94] J.L. Gauvain and Ch.H. Lee. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEEISA*, 2(2), 1994.
- [GM85] Milligan G.W. and Cooper M.C. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):pp. 159–179, 1985.
- [Gre86] S.D. Greenwald. *Development and analysis of a ventricular fibrillation detector*. PhD thesis, MIT Dept. of Electrical Engineering and Computer Science, 1986.

- [Gre90] S.D. Greenwald. *Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information*. PhD thesis, Harvard-MIT Division of Health Sciences and Technology, 1990.
- [GS99] S. Gaffney and P. Smyth. Trajectory clustering with mixtures of regression models. Technical Report 99-15, University of California, 1999.
- [GT74] R.C. González and J.T. Tou. *Pattern Recognition Principles*. Addison-Wesley Publishing Company, 1974.
- [HAA96] Y. Hamam and T. Al-Ani. Simulated annealing approach for training hidden markov models. In *Working Conference on Optimization-Based Computer-Aided Modeling and Design*, ESIEE, France, 1996.
- [Hal67] A.V. Hall. Methods for demonstrating resemblance in taxonomy and ecology. *Nature*, 214:830–831, 1967.
- [Har74] J.A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, 1974.
- [Har90] H. Harold. *Diagnóstico Electrocardiográfico y Vectorcardiográfico*. Salvat, 1990.
- [HK01] J. Han and M. Kamber. *Data Mining*. Morgan Kaufmann Publishers, 2001.
- [HKT01] J. Han, M. Kamber, and A.K. Tung. *Spatial clustering methods in data mining: A survey*. Geographic Data Mining and Knowledge Discovery. Taylor and Francis, 2001.
- [HO00] A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000.
- [Hof97] T. Hofmann. Pairwise data clustering by deterministic annealing. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 19(1):1–14, 1997.

- [Hor75] S.L. Horowitz. A syntactic algorithm for peak detection in waveforms with applications to cardiography. *Communications ACM*, 18:281–285, 1975.
- [IM98] H. Inoue and A. Miyazaki. A noise reduction method for ecg signals using the dyadic wavelet transform. *IEICE Trans. Fundamentals*, E81-A(6):pp. 1001–1007, June 1998.
- [Jai88] A. Jain. *Algorithms for Clustering Data*. Prentice Hall, 1988.
- [JDM00] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 2000.
- [Jel76] F. Jelinek. Continuous speech recognition by statistical methods. *Proceedings of the IEEE*, 64(4):532–536, 1976.
- [Jel99] F. Jelinek. *Statistical methods for speech recognition*. MIT Press, 1999.
- [JGS97] M.I. Jordan, Z. Ghahramani, and L. Saul. Hidden markov decision trees. *Advances in Neural Information Processing Systems*, 1997.
- [JJ94] M.I. Jordan and R. Jacobs. Hierarchical mixtures of experts and the em algorithm. *Neural Computation*, 6:181–214, 1994.
- [JLTC92] R. Jane, P. Laguna, N.V. Thakor, and P. Caminal. Adaptive baseline wander removal in the ecg: Comparative analysis with cubic splines technique. *Computers in Cardiology, IEEE Computer Society Press*, pages pp. 143–146, 1992.
- [JMF99] A.K. Jain, M.N. Murty, and P.J. Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3), 1999.
- [JMT⁺96] F. Jager, G.B. Moody, A. Taddei, G. Antolic, M. Zabukovec, M. Skrjanc, M. Emdin, and R.G. Mark. Development of a long-term database for assessing the performance of transient ischemia detectors. *Computers in Cardiology*, pages pp. 481–484, 1996.

- [Joh04] V. Johneff. Complex valued wavelet analysis for qrs detection in ecg signals. *Bioinformatics, Images, and Wavelets*, pages 134–136, 2004.
- [JR90] B.H. Juang and L.R. Rabiner. The segmental k-means algorithm for estimating the parameters of hidden markov models. *IEEE Transaction on Accoustic, Speech and Signal Processing*, 38(9):1639–1641, 1990.
- [JT93] Boberg J. and Salakoski T. General formulation and evaluation of agglomerative clustering methods with metric and non-metric distances. *Pattern Recognition*, 26(9):1395–1406, 1993.
- [JTM⁺03] F. Jager, A. Taddei, G.B. Moody, M. Emdin, G. Antolic, R. Dorn, A. Smrdel, C. Marchesi, , and R.G. Mark. Long-term st database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia. *Medical and Biological Engineering and Computing*, 41(2):pp. 172–183, 2003.
- [Jua99] A. Juan. *Optimización de Prestaciones en Técnicas de Aprendizaje No Supervisado y su Aplicación al Reconocimiento de Formas*. PhD thesis, Universidad Politécnica de Valencia, 1999.
- [KB98] S.P. Krozaitis and A.A. Basuhail. Adaptive wavelet threshold selection using higher-order statistics for signal denoising. *SPIE*, vol. 3391:pp. 68–74, 1998.
- [KCMT01] S. Kwong, C.W. Chau, K.F. Man, and K.S Tang. Optimisation of hmm topology and its model parameters by genetic algorithms. *Pattern Recognition*, 34:509–522, 2001.
- [Ke94] A. Krogh and et al. Hidden markov models in computational biology: application to protein modleing. *Journal of Molecular Biology*, 235(1501–1531), 1994.
- [KJ94] D. Kilpatrick and P.R. Johnston. Origin of the electrocardiogram. *IEEE Engineering in Medicine and Biology*, August–September:479–486, 1994.

- [KJ96] A. Koski and M. Juhola. Segmentation of digital signals based on estimated compression ratio. *IEEE transactions on Biomedical Engineering*, 43(9), 1996.
- [KJM95] A. Koski, M. Juhola, and M. Mersite. Syntactic recognition of ecg signals by attributed finite automata. *Pattern Recognition*, 28(12):pp. 1927–1940, 1995.
- [KK96] R. Krishnapuram and J.M. Keller. The possibilistic c-means algorithm: Insights and recommendations. *IEEE Transactions on Fuzzy Systems*, 4(3):385–393, 1996.
- [KMBB99] S. Kadambe, R. Murray, and G. Boudreaux-Bartels. Wavelet transform-based qrs complex detector. *Biomedical Engineering Trans.*, 46(7):838–848, 1999.
- [KNF95] W.L. Koontsz, P.M. Narendra, and K. Fukunaga. Branch and bound clustering algorithm. *IEEE Transaction on Computers*, 24(9), 1995.
- [Koh90] T. Kohonen. Self-organized formation of topologically correct feature maps. In J. W. Shavlik and T. G. Dietterich, editors, *Readings in Machine Learning*. Morgan Kaufmann Publishers, Inc., 1990.
- [Kol] E. Kolatch. Clustering algorithms for spatial databases: A survey.
- [Kos96] A. Koski. Modelling ecg signals with hidden markov models. *Artificial Intelligence in Medicine*, 8:453–471, 1996.
- [Kou98] K. Koutroumbas. Neural network architecture for selecting the maximum input. *International Journal of Computer Mathematics*, 68(1–2), 1998.
- [KW98] K. Berkner and R.O. Wells. Wavelet transforms and denoising algorithms. *IEEE*, pages pp. 1639–1643, 1998.
- [Laf95] J.D. Laffery. Gibbs-markov models. In *Computing Science and Statistics: Proceedings of the 27th Symposium on the Interface*, 1995.

- [Lag90] P. Laguna. *Nuevas técnicas de procesado de señales electrocardiográficas: aplicación a registros de larga duración*. PhD thesis, Instituto de Cibernética de Cataluña, 1990.
- [LB00] C. Li and G. Biswas. A bayesian approach to temporal data clustering using hidden markov models. *International Conference on Machine Learning*, pages 543–550, 2000.
- [LBBC01] R. Lepage, J.M. Boucher, J.J. Blanc, and J.Ch. Cornilly. Ecg segmentation and p-wave feature extraction: Application to patients prone to atrial fibrillation. In *23rd Annual Conference IEEE/EMBS*, 2001.
- [Lip87] R.P. Lippmann. An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4(2), 1987.
- [LMGM97] P. Laguna, R.G. Mark, A. Goldberger, and G.B. Moody. A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg. *Computers in Cardiology*, 24:pp. 673–676, 1997.
- [LZT95] C. Li, C. Zheng, and C. Tai. Detection of ecg characteristic points using wavelet transforms. *IEEE Transactions on Biomedical Engineering*, 42(1):21–28, 1995.
- [Mac97] D.J.C. MacKay. Ensemble learning for hidden markov models, 1997. <http://www.inference.phy.cam.ac.uk/mackay/abstracts/ensemblePaper.html>.
- [Mal91] S. Mallat. Zero-crossings of a wavelet transform. *IEEE Transactions on Information Theory*, 37:1019–1033, 1991.
- [Mat01] The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA, USA. *Statistics toolbox*, 5th edition, May 2001.
- [MC02] P. Micó and D. Cuesta. Early dyslexia detection techniques by means of oculographic signals. In *European Medical and Biological Engineering Conference*, 2002.

- [MCN04a] P. Micó, D. Cuesta, and D. Novák. High-speed feature extraction in holter electrocardiogram using principal component analysis. In *The 17th international EURASIP conference BIOSIGNAL 2004*, 2004.
- [MCN04b] P. Micó, D. Cuesta, and D. Novák. Pre-clustering of electrocardiographic signals using ergodic hidden markov models. *International Conference on Statistical and Structural Pattern Recognition*, LNCS 3138:939–947, 2004.
- [MCN05a] P. Micó, D. Cuesta, and D. Novák. Clustering improvement for electrocardiographic signals. *International Conference on Image Analysis and Processing*, LNCS 3617:892–899, 2005.
- [MCN05b] P. Micó, D. Cuesta, and D. Novák. Polygonal approximation of holter registers: A comparative study for electrocardiographic signals time compression. *Computational Intelligence in Medicine and Healthcare*, pages 323–329, 2005.
- [MH79] D. Michael and J. Houchin. Automatic eeg analysis: A segmentation procedure based on autocorrelation function. *Electroencephalogr., Clin. Neurophysiol.*, 46:pp. 232, 1979.
- [MJ96] M. Meila and M.I. Jordan. Learning finite motion by markov mixtures of experts. *Advances in Neural Information Processing Systems*, 8:1003–1009, 1996.
- [MK97a] G.J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, 1997.
- [MK97b] C.R. Meyer and H.N. Keiser. Electrocardiogram baseline noise estimations and removal using cubic splines and state-space computation techniques. *Computers and Biomedical Research*, pages pp. 459–470, 1997.
- [MM83] G.B. Moody and R.G. Mark. A new method for detecting atrial fibrillation using r-r intervals. *Computers in Cardiology*, 10:pp. 227–230, 1983.

- [MM89] G. Moody and R. Mark. Qrs morphology representation and noise estimation using the karhunen-loeve transform. *Computers in Cardiology*, 1989.
- [MM98] R. Mark and G. Moody. Mit-bih arrhythmia data base directory. Massachusetts Institute of Technology, 1998.
- [MMG88] G.B. Moody, R.G. Mark, and A.L. Goldberger. Evaluation of the trim ecg data compressor. *Computers in Cardiology*, 15:pp. 167–170, 1988.
- [MMM84] G.B. Moody, W.E. Muldrow, and R.G. Mark. A noise stress test for arrhythmia detectors. *Computers in Cardiology*, 1984.
- [Moe00] Perry Moerland. *Mixture models for Unsupervised and Supervised Learning*. PhD thesis, Computer Science Department of the Swiss Federal Institute of technology, Lausanne, 2000.
- [MP00] G.J. McLachlan and D. Peel. *Finite Mixture Models*. John Wiley & Sons, 2000.
- [MPC⁺81] C.N. Mead, H.R. Pull, J.S. Cheng, K.W. Clark, and L.J. Thomas. A frequency domain based qrs classification algorithm. *Computers in Cardiology*, pages pp. 351–354, 1981.
- [MPH⁺02] J.E. Mietus, C-K Peng, I. Henry, R.L. Goldsmith, and A.L. Goldberger. The pnnx files: re-examining a widely used heart rate variability measure. *Heart*, 2002.
- [Mur02] K.P. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, 2002.
- [MYD96] C. Morimoto, Y. Yaccob, and L. Davis. Recognition of head gestures using hidden markov models. Technical report, University of Maryland, 1996.
- [MZ00] Iain L. MacDonald and Walter Zucchini. *Hidden Markov and other Models for Discrete-valued Time Series*. Chapman & Hall/CRC, 2000.

- [NAaH⁺04] D. Novák, T. Al-ani, Y. Hamam, D. Cuesta, P. Micó, and L. Lhotská. Unsupervised learning of holter ecg signals using hmm optimized by simulated annealing. In *The 17th international EURASIP conference BIOSIGNAL 2004*, 2004.
- [NCAa⁺04] D. Novák, D. Cuesta, T. Al-ani, M. Aboy, P. Micó, and L. Lhotská. Speech recognition methods applied to biomedical signals processing. 26th Annual International Conference IEEE Engineering in Medicine and Biology Society, September 2004.
- [NCAa⁺05] D. Novák, D. Cuesta, T. Al-ani, A. Hamam, M. Aboy, P. Micó, and L. Lhotská. Clustering of biological signals using hidden markov models. IEEE - Reviewing, 2005.
- [NCE⁺00] D. Novák, D. Cuesta, V. Eck, J. C. Perez, and G. Andreu. Denoising electrocardiogram signal using adaptive wavelets. In *The 15th international EURASIP conference BIOSIGNAL 2000*, pages pp. 18–20, 2000.
- [NCML03] D. Novák, D. Cuesta, P. Micó, and L. Lhotská. Number of arrhythmia beats in holter ecg: how many clusters? In *25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2003.
- [NCML04] D. Novák, D. Cuesta, P. Micó, and L. Lhotská. Morphology analysis of biological signals using hidden markov models. 17th International Conference on Pattern Recognition, August 2004.
- [Nov02] D. Novák. Wavelet image denoising-a comparative study. Technical report, Department of Cybernetics, Czech Technical University in Prague, 2002.
- [Obe01] B. Obermaier. Hidden markov models for online classification of single trial eeg data. *Pattern Recognition Letters*, 22(1299–1309), 2001.
- [ODK96] M. Ostendorf, V. Digalakis, and O. Kimball. From hmm's to segmented models: a unied view of stochastic modelling for speech recognition. *IEEE Transaction on Speech and Audio Processing*, 4(5):360–378, 1996.

- [OFC99] T. Oates, L. Firoiu, and P.R. Cohen. Clustering time series with hidden markov models and dynamic time warping. In *Workshop on Sequence Learning, IJCAI-99*, pages 17–21, 1999.
- [OGJL97] S. Olmos, J. García, R. Jané, and P. Laguna. Análisis de la señal ecg con expansiones ortogonales de reducido número de coeficientes. *Comunicaciones del XV Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, 1997.
- [OIEC95] N.J. Outram, E.C. Ifeachor, P.W. Van Eetvelt, and J.S.H. Curnow. Techniques for optimal enhancement and feature extraction of fetal electrocardiogram. *IEEE Proceeding*, 142(6):pp. 482–489, 1995.
- [Oka79] M. Okada. A digital filter for the qrs detection. *IEEE Transactions on Biomedical Engineering*, 26(12):pp. 700–703, 1979.
- [OL97] B.J. Oommen and R.K. Loke. Pattern recognition of strings with substitutions. *Pattern Recognition*, 30(5):789–800, 1997.
- [OS89] A.V. Oppenheim and R.W. Schaffer. *Discrete-Time Signal Processing*. Prentice-Hall International, 1989.
- [Par99] M.K. Park. *Cardiología pediátrica*. Harcourt Brace, 1999.
- [PD00] E. Pranckeviciénė and L. Druciunas. Classification of activity of epileptic eeg. In *Proceedings of the European BIOSIGNAL 2000*, pages 168–169, 2000.
- [PH74] T. Pavlidis and S.L. Horowitz. Segmentation of plane curves. *IEEE Trans. Comput.*, C-23(8):pp. 860–870, 1974.
- [PJ92] W. Philips and G.D. Jonghe. Data compression of ecg's by high-degree polynomial approximation. *IEEE Trans. Biomed Eng.*, 39:pp. 330–337, 1992.
- [PMM02] A. Panuccio, M. Bicego, and V. Murino. A hidden markov model-based approach to sequential data clustering. In T. Caelli, A. Amin, R.P.W. Duin, M. Kamel, , and D. de Ridder Eds, editors, *Structural, Syntactic and Statistical Pattern Recognition, LNCS 2396*, pages 734–742. Springer, 2002.

- [PR98] W.D. Penny and S.J. Roberts. Gaussian observations hidden markov models for eeg analysis. Technical report, Imperial College of Science, 1998.
- [PT95] Soo-Chang Pei and Chien-Cheng Tseng. Elimination of ac interference in electrocardiogram using iir notch filter with transient suppression. *IEEE Transactions on Biomedical Engineering*, 42(11):pp. 1128–1132, 1995.
- [PTCJ90] P.Laguna, N.V. Thakor, P. Caminal, and R. Jane. Low-pass differentiators for biological signals with known spectra: Application to eeg signal processing. *IEEE Trans. Biomedical Eng.*, 37(4):pp. 420–424, 1990.
- [Rab89] R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77, 1989.
- [Ram72] U. Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1:244–257, 1972.
- [Ran02] R.M. Rangayyan. *Biomedical Signal Analysis. A Case-Study Approach*. Wiley-IEEE Press, 2002.
- [RB94] G. Radons and J.D. Becker. Analysis, classification, and coding of multielectrode spike trains with hidden markov models. *Biological Cybernetics*, 71:359–373, 1994.
- [RJ74] Calinski R.B. and Harabasz J. A dendrite method for cluster analysis. *Communication in Statistics*, 3:pp. 1–27, 1974.
- [RMR02] I. Rezek, M.Gibbs, and S.J. Roberts. Maximum a posteriori estimation of coupled hidden markov models. *Journal of VLSI Signal Processing*, 32:55–66, 2002.
- [RPR00] I. Rezek, P.Sykacek, and S.J. Roberts. Learning interaction dynamics with coupled hidden markov models. *IEE Special Issue Proceedings Science, Measurement and Technology*, 147(6):345–350, 2000.

- [RR92] B.K. Ray and K.S. Ray. An algorithm for polygonal approximation of digitised curves. *Pattern Recognition Letters*, 13:pp. 489–496, 1992.
- [RR02] I. Rezek and S.J. Roberts. Learning ensemble hidden markov models for biosignal analysis. In *14th International Conference on Digital Signal Processing*, Greece, 2002.
- [RV03] Y.V. Ramana Rao and N. Venkateswaran. Allpass lattice structure based second order digital iir notch filter for removing dc and very low frequencies. *IEEE Aerospace Conference*, 2003.
- [RZ86] D.E Rumelhart and D. Zipser. Feature discovery by competitive learning. *Cognitive Science*, 9:75–112, 1986.
- [SBNP81] L. Sörnmo, P.O. Börjesson, M.Ñygarads, and O. Pahlm. A method for evaluation of qrs shape features using a mathematical model for the ecg. *IEEE Transactions on Biomedical Engineering*, 28(10), 1981.
- [SC94] J. Stutz and P. Cheeseman. Autoclass - a bayesian approach to classification. In J.Skilling and S. Sibisi, editors, *Maximum Entropy and Bayesian Methods*. Kluwer Academic Publishers, 1994.
- [Sch97] P. Scheunders. A gennetic c-means clustering algorithm applied to color image quantization. *Pattern Recognition*, 30(6):859–866, 1997.
- [Sco02] Steven L. Scott. Bayesian methods for hidden markov models. *Journal of the American Statistical Association*, 97:337–351, 2002.
- [SCSA04a] J.J. Segura, D. Cuesta, L. Samblás, and M. Aboy. A microcontroller-based portable electrocardiograph recorder. *IEEE Transactions on Biomedical Engineering*, 51(9):pp. 1686–1690, 2004.
- [SCSA04b] J.J. Segura, D. Cuesta, L. Samblás, and M. Aboy. A microcontroller-based portable electrocardiograph recorder. *IEEE Transactions on Biomedical Engineering*, 51(9):1686–1690, 2004.

- [SG80] J. Sklansky and V. Gonzales. Fast polygonal approximation of digitised curves. *Pattern Recognition*, 12:pp. 327–331, 1980.
- [SJ95] L. Saul and M.I. Jordan. Boltzmann chains and hidden markov models. *Advances in Neural Information Processing Systems*, 7, 1995.
- [SKK76] G. Stockman, L. Kanal, and M.C. Kyle. Structural pattern recognition of carotid pulse waves using a general waveform parsing system. *Communicat. ACM*, 19(12):688–695, 1976.
- [SMLZ98] E. Sanchez, J. Millet, J.J. López, and C. Zywietz. Comparativa en la discriminación de ritmos normales y anormales de ecg estándar por medio del análisis lineal discriminante y Árboles de decisión. *Comunicaciones del XVI Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, pages pp. 151–154, 1998.
- [Smy94] P. Smyth. Hidden markov models for fault detection in dynamic systems. *Pattern Recognition*, 27(1):149–164, 1994.
- [Spa80] H. Spath. *Cluster Analysis Algorithms*. Ellis Horwood, 1980.
- [Sör93] L. Sörnmo. Time-varying digital filtering of ecg baseline wander. *Medical and Biological Engineering and Computers*, pages pp. 503–508, 1993.
- [SR99] J.C. Stiller and G. Radons. Online estimation of hidden markov models. *IEEE Signal Processing Letters*, 6(8):213–215, 1999.
- [SS94] S. Suppappola and Y. Sun. Nonlinear transforms of ecg signals for digital qrs detection: A quantitative analysis. *IEEE Transactions on Biomedical Engineering*, 41(4):pp. 397–400, 1994.
- [SSL96] B. Simón, L. Sörnmo, and P. Laguna. Mejora en el alineamiento de registros ecg usando interpolación. *Comunicaciones del XIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, pages pp. 49–51, 1996.
- [SW96] Y. Singer and M.K. Warmuth. Training algorithms for hidden markov models using entropy based distance functions. In *Ad-*

- vances in Neural Information Processing Systems 9*, pages 641–647. Morgan Kaufmann Publishers, 1996.
- [Tan58] T. Tanimoto. An elementary mathematical theory of classification and prediction. Technical report, IBM Corp., 1958.
- [TB99] M.E. Tipping and C.M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society*, pages 611–622, 1999.
- [TK99] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 1999.
- [TLF99] S. Thrun, J. Langford, and D. Fox. Monte carlo hidden markov models: Learning non-parametric models of partially observable stochastic processes. In *Proceedings of the International Conference on Machine Learning*, Bled, Slovenia, 1999.
- [TLJ⁺01] M.C. Teich, S.B. Lowen, B.M. Jost, K. Vibe-Rhymer, and C. Heneghan. Heart rate variability: measures and models. *Non-linear Biomedical Signal Processing*, II(6):pp. 159–213, 2001.
- [TS89] P. Trahanias and E. Skordalakis. Bottom-up approach to the ecg pattern recognition problem. *Medical and Biological Engineering and Computing*, pages pp. 221–229, 1989.
- [TS90] P. Trahanias and E. Skordalakis. Syntactic pattern recognition of the ecg. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):pp. 648–656, 1990.
- [TW99] D. Tran and M. Wagner. Fuzzy hmms for speech and speaker recognition. In *The 18th International Conference of the North American Fuzzy Information Society, BAFIPS'99, USA, 1999*.
- [TW00] D. Tran and M. Wagner. Frame-level hidden markov models. In M. Mohammadian, editor, *Advances in Intelligent Systems: Theory and Applications*, pages 252–259. IOS Press, 2000.
- [TWT84] N.V. Thakor, J.G. Webster, and W.J. Tompkins. Estimation of qrs complex power spectra for design of a qrs filter. *IEEE Trans-*

- actions on Biomedical Engineering*, 31(11):pp. 702–706, November 1984.
- [VFLC95] V. Di Virgilio, C. Francalancia, S. Lino, and S. Cerutti. Ecg fiducial points detection through wavelet transform. *IEEE/EMBC and CMBEC*, pages pp. 1051–1052, 1995.
- [VGL97] E. Vullings, J. García, and P. Laguna. Waveform detection in holter ecg using dynamic time warping. *Comunicaciones del XV Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, pages pp. 313–316, 1997.
- [Vit67] A.J. Viterbi. Error bounds for convolutional codes and an asymptotical optimal decoding algorithm. *IEEE Transaction on Information Theory*, IT(13):260–269, 1967.
- [VK92] J.A. Vlontzos and S.Y. Kung. Hidden markov models for character recognition. *IEEE Transaction on Image Processing*, 1(4):539–543, 1992.
- [WS98] Andreas S. Weigend and Shanming Shi. Predicting daily probability distributions of s&p500 returns. Technical report, Stanford University, Psychology Department, 1998.
- [YHS97] M. Yang, W. Hu, and L. Shyu. Ecg events detection and classification using wavelet and neural networks. *Proceedings 19th International Conference IEEE/EMBS*, pages pp. 280–281, 1997.
- [YXC97] J. Yang, Y. Xu, and Ch.S. Chen. Human action learning via hidden markov model. *IEEE Transaction on Systems Man and Cybernetics. Part A: Systems and Humans*, 27(1), 1997.

Lista de acrónimos

AHA	<i>American Heart Association</i>
ARHMM	<i>Auto-Regressive Hidden Markov Model</i>
BCS	<i>Between Cluster Similarity</i>
BIC	<i>Bayesian Inference Criterion</i>
BIH	<i>Beth Israel Hospital</i>
BMM	<i>Buried Markov Model</i>
CHMM	<i>Coupled Hidden Markov Model</i>
CPU	<i>Central Process Unit</i>
DCMM	<i>Double Chain Markov Model</i>
DTW	<i>Dynamic Time Warping</i>
DWT	<i>Discrete Wavelet Transform</i>
ECG	<i>Electrocardiograma</i>
EEG	<i>Electroencefalograma</i>
EM	<i>Expectation-Maximization</i>
EMBS	<i>Engineering in Medicine and Biology Society</i>
FFT	<i>Fast Fourier Transform</i>
FHMM	<i>Factorial Hidden Markov Model</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
FIR	<i>Finite Impulse Response</i>
GA	<i>Genetic Algorithms</i>
GMM	<i>Gaussian Mixture Model</i>
HHMM	<i>Hierarchical Hidden Markov Model</i>
HMM	<i>Hidden Markov Model</i>
HMMDT	<i>Hidden Markov Model Decision Trees</i>

IBM	<i>International Business Machines</i>
IDWT	<i>Inverse Discrete Wavelet Transform</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
IIR	<i>Infinite Impulse Response</i>
IOHMM	<i>Input-Output Hidden Markov Models</i>
LMS	<i>Least Mean Squared</i>
lpm	<i>Latidos Por Minuto</i>
MAE	<i>Mean Absolute Error</i>
MANOVA	<i>Multivariate ANalysis Of VAriance</i>
MAP	<i>Maximum A Posteriori</i>
MCHMM	<i>Monte-Carlo Hidden Markov Models</i>
MD	<i>Medida de Disimilitud</i>
MIT	<i>Massachusetts Institute of Technology</i>
ML	<i>Maximum Likelihood</i>
MLP	<i>Multi-Layer Perceptron</i>
MMDL	<i>Mixture Minimum Description Length</i>
MRF	<i>Markov Random Field</i>
MS	<i>Medida de Similitud</i>
MSE	<i>Mean Squared Error</i>
NN	<i>Neural Network</i>
PCA	<i>Principal Component Analysis</i>
PPCA	<i>Probabilistic Principal Component Analysis</i>
RAM	<i>Random Access Memory</i>
SHMM	<i>Segmented Hidden Markov Model</i>
SNR	<i>Signal to Noise Relation</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
VAR	<i>Variational Learning Estimation</i>
VxC	<i>Visión por Computador</i>