COMBINACIÓN DE PPCA Y HMM PARA LA IDENTIFICACIÓN DE INFARTO AGUDO DE MIOCARDIO

MAURICIO ALVAREZ
Ingeniero Electrónico, Ms.Eng.
Profesor Auxiliar
Grupo de Investigación Control e Instrumentación
Universidad Tecnológica de Pereira
malvarez@ohm.utp.edu.co

RICARDO HENAO
Ingeniero Electrónico, Ms.Eng.
Profesor Auxiliar
Grupo de Investigación "LIDER"
Universidad Tecnológica de Pereira
rhenao@utp.edu.co

RESUMEN
Este artículo presenta una nueva versión de análisis de componentes principales que utiliza una cadena oculta de Markov con la finalidad de obtener una representación optimizada de datos observados a través del tiempo para la identificación de infarto agudo de miocardio. La novedad del método propuesto consiste principalmente en la manera en la cual una técnica de reducción de dimensionalidad estática ha sido combinada con un modelo clásico de mezclas en el tiempo, para mejorar las capacidades de transformación, reducción y clasificación de series de tiempo. Los resultados experimentales muestran mejoras substanciales en la precisión de clasificación de registros con infarto agudo de miocardio, incluso con representaciones altamente reducidas de las series multivariadas transformadas.

PALABRAS CLAVES: Análisis de componentes principales, Análisis probabilístico de componentes principales, Modelos ocultos de Markov, Clasificación de series de tiempo, Infarto agudo de miocardio.

ABSTRACT
This paper introduces a new temporal version of PCA by using a hidden Markov model in order to obtain optimized representations of observed data through time for acute myocardial infarction identification. The novelty of the proposed method consists mainly in the way in which static dimensionality reduction techniques have been combined with a classic mixture model in time, to enhance the capabilities of transformation, reduction and classification of time series, in this case, ECG records. Experimental results show improvements in classification accuracies over records with acute myocardial infarction even with highly reduced representations of the transformed multivariate series.

KEYWORDS: Principal component analysis, Probabilistic principal component analysis, Hidden Markov models, Time series classification, Acute myocardial infarction.

1. INTRODUCCIÓN

El Análisis de Componentes Principales (PCA) es una técnica popular empleada para la extracción de características, reducción de dimensionalidad y probablemente la más utilizada de las técnicas de análisis multivariado [1]. Una de las definiciones más comunes de PCA es que dado un conjunto *d*-dimensional de N vectores observados $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$, los q ejes principales $\mathbf{W} = \{\mathbf{w}_1, ..., \mathbf{w}_p\}$ son aquellos ejes ortonormales en los cuales la varianza retenida sobre una proyección lineal es máxima.

Sin embargo, PCA posee una desventaja que consiste en la falta de función de densidad de probabilidad o modelo generativo asociado. El Análisis Probabilístico de Componentes Principales (PPCA) [2], elimina dicha desventaja considerando un modelo estadístico de variables latentes con estructura de ruido isotrópico, de modo que PCA queda inmerso implícitamente en una etapa de aprendizaje de parámetros que utiliza el método de estimación de máxima verosimilitud (ML).

La utilización de PPCA como técnica de reducción de dimensionalidad global posee varios inconvenientes [3], por ejemplo, la necesidad de un conjunto de datos suficientemente grande para ajustar el igualmente grande conjunto de parámetros del modelo. Sin embargo, las técnicas de reducción de dimensionalidad local, mejoran la capacidad de reducción de dimensionalidad mediante la combinación de modelos locales más simples, es decir, con menos parámetros por modelo. Una manera natural de obtener representaciones locales y a la vez reducción de la dimensión de los datos es utilizar mezclas de modelos de variable latente. Si el contexto es temporal, una cadena oculta de Markov (HMM) podría ser un modelo apropiado para dicha tarea. Si bien las mezclas de PPCA ya han sido formuladas [4], el contexto no es temporal.

En este artículo, una versión temporal de PPCA es presentada, utilizando una HMM como medio para

Fecha de Recepción: 31 Agosto de 2006
Fecha de Aceptación: 20 Noviembre de 2006

obtener una representación optimizada de los datos observados a través del tiempo. Desde este punto de vista, cada observación posee una representación local correspondiente al estado más probable producido por una HMM entrenada. La integración entre PPCA y HMM se realizó incorporando una factorización de los datos de entrenamiento y los pesos de responsabilidad en la matriz de covarianza del modelo de observación de la HMM de manera que PPCA sea calculado fácilmente.

De otro lado, el síndrome coronario agudo (SCA), se refiere a diferentes situaciones clínicas secundarias a la obstrucción brusca del flujo coronario, de gravedad y pronóstico variable. El infarto agudo de miocardio (IAM) es la más grave de esas situaciones y se produce por la oclusión aguda de una o varias arterias coronarias con la consiguiente necrosis miocardíaca [5]. Esta es la manifestación más importante de la cardiopatía isquémica que, junto al resto de enfermedades cardiovasculares, sigue siendo una de las primeras causas de muerte en la población colombiana [6]. En este sentido, el análisis de las señales de ECG que permita identificar automáticamente la presencia de actividad anormal (infarto agudo de miocardio), es una herramienta de gran ayuda para los sistemas de soporte al diagnóstico de patologías cardíacas.

Este artículo está organizado como sigue: las secciones 2 y 3 contienen revisiones de la teoría de PPCA y HMM respectivamente. En la sección 4, se presenta la extensión de PPCA a través del tiempo. En la sección 5 se reportan los resultados experimentales obtenidos para bases de datos de registros de ECG utilizados para la identificación de infarto agudo de miocardio, y finalmente en la sección 6 las conclusiones y discusión del trabajo.

2. ANÁLISIS PROBABILÍSTICO DE COMPONENTES PRINCIPALES

Un modelo de variables latentes pretende relacionar un vector observado $\mathbf{x}$ de dimensión d con un vector de variables latentes $\mathbf{y}$ de dimensión q. Partiendo de la perspectiva de Análisis Factorial (FA) [7], $\mathbf{x}$ puede ser expresado como una combinación lineal de vectores base y ruido,

$$\mathbf{x} = \mathbf{W}\mathbf{y} + \boldsymbol{\mu} + \boldsymbol{\varepsilon} \qquad (1)$$

donde la matriz $\mathbf{W}$ de dimensiones $d \times q$, es una base que relaciona ambos conjuntos de variables, $\boldsymbol{\varepsilon}$ es un proceso de ruido no dependiente de $\mathbf{y}$, y el vector $\boldsymbol{\mu}$ permite que el modelo tenga media diferente de cero. Convencionalmente, las variables latentes están definidas como independientes y gaussianas, $p(\mathbf{y}) \sim \mathcal{N}(0, \mathbf{I})$ y $p(\boldsymbol{\varepsilon}) \sim \mathcal{N}(0, \boldsymbol{\Psi})$, con $\boldsymbol{\Psi}$ diagonal y $\mathbf{W}$ la matriz de parámetros que contiene los pesos de los factores.

La ecuación (1) induce una distribución gaussiana para las observaciones como,

$$p(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$$

donde $\mathbf{C} = \mathbf{W}\mathbf{W}' + \boldsymbol{\Psi}$. Para obtener el modelo de PCA, $\boldsymbol{\Psi}$ debe ser isotrópica en vez de diagonal con varianzas residuales dadas por $\boldsymbol{\Psi}_{ii} = \sigma^2$ de manera que el modelo de ruido es asumido como $\boldsymbol{\Psi} = \sigma^2\mathbf{I}$ y $\mathbf{C} = \mathbf{W}\mathbf{W}' + \sigma^2\mathbf{I}$, con esto, el modelo de PCA puede ser obtenido mediante procedimientos iterativos utilizando máxima verosimilitud, debido a que no hay solución analítica en forma compacta.

Las estimaciones para $\mathbf{W}$ y σ^2 se pueden obtener maximizando iterativamente $\mathcal{L} = \log\left(p(\mathbf{x}\,|\,\mathbf{W}, \boldsymbol{\mu}, \sigma^2)\right)$,

$$\mathcal{L} = -\frac{N}{2}\left\{d\log(2\pi) + \log|\mathbf{C}| + \operatorname{tr}\left(\mathbf{C}^{-1}\mathbf{S}\right)\right\} \qquad (2)$$

donde

$$\mathbf{S} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})' \qquad (3)$$

El estimador para $\boldsymbol{\mu}$ es la media muestral, esto es,

$$\boldsymbol{\mu}_{ML} = \overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$$

Se puede demostrar [2], que la maximización de los parámetros en (2) está dada por,

$$\widehat{\mathbf{W}} = \mathbf{S}\mathbf{W}(\sigma^2\mathbf{I} + \mathbf{M}^{-1}\mathbf{W}'\mathbf{S}\mathbf{W})^{-1} \qquad (4)$$

$$\hat{\sigma}^2 = \frac{1}{d}\operatorname{tr}(\mathbf{S} - \mathbf{S}\mathbf{M}^{-1}\widehat{\mathbf{W}}') \qquad (5)$$

donde $\widehat{\mathbf{W}}$, significa "nuevo" y $\mathbf{M} = \mathbf{W}'\mathbf{W} + \sigma^2\mathbf{I}$. Con esto, las ecuaciones en (4) y (5) son iteradas en secuencia hasta que se juzga que el algoritmo ha convergido.

3. MODELOS OCULTOS DE MARKOV

Un modelo oculto de Markov es básicamente una cadena de Markov donde las observaciones en la salida son una variable aleatoria generada conforme a una función de probabilidad asociada a cada estado [8]. Formalmente, un modelo oculto de Markov de N_s estados está definido por

- $\mathbf{A} = \{a_{ij}\}$. Una matriz de transición de probabilidad donde a_{ij} denota la probabilidad de hacer la transición del estado i al estado j, esto es,

$$a_{ij} = P(s_n = j \mid s_{n-1} = i)$$

- $B = \{b_i(k)\}$. Una matriz de probabilidad de salida, donde $b_i(k)$ es la probabilidad de emitir el símbolo o_k cuando se procede del estado i. Sea $X = \{\mathbf{x}_n\}$ la salida observada de la HMM. La secuencia de estados $S = s_1, s_2, ..., s_n, ...$ no es observada y $b_i(k)$ puede ser escrita como sigue

$$b_i(k) = P(\mathbf{x}_n = o_k \mid s_n = i)$$

Si la observación no proviene de un conjunto finito sino de un espacio continuo, la distribución discreta de la salida puede ser reemplazada por una función de densidad de probabilidad continua $b_i(\mathbf{x}_n)$. De entre varias alternativas, una mezcla multivariada de funciones de densidad gaussianas es usualmente empleada.

- $\boldsymbol{\pi} = \{\pi_i\}$. Una distribución inicial para los estados donde

$$\pi_i = P(s_0 = i) \quad 1 \le i \le N_s$$

Por simplicidad, el conjunto de parámetros es denotado como

$$\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$$

Para entrenar los HMM, se emplea el algoritmo estándar de maximización de la esperanza (EM) [7]. En particular, cuando el modelo de observación es una mezcla de gaussianas,

$$p_\lambda(\mathbf{x}_n \mid s_n = j) = \sum_{m=1}^{M} c_{jm} \mathcal{N}(\mathbf{x}_n; \boldsymbol{\mu}_{jm}, \mathbf{S}_{jm}) \quad (6)$$

donde M, es el número de componentes en la mezcla. Se puede demostrar que las fórmulas de reestimación están dadas por

$$\begin{aligned}
\hat{c}_{jk} &= \frac{\sum_{n=1}^{N} \gamma_n(j,k)}{\sum_{n=1}^{N} \sum_{k=1}^{M} \gamma_n(j,k)} \\
\hat{\boldsymbol{\mu}}_{jk} &= \frac{\sum_{n=1}^{N} \gamma_n(j,k) \mathbf{x}_n}{\sum_{n=1}^{N} \gamma_n(j,k)} \\
\hat{\mathbf{S}}_{jk} &= \frac{\sum_{n=1}^{N} \gamma_n(j,k)(\mathbf{x}_n - \hat{\boldsymbol{\mu}}_{jk})(\mathbf{x}_n - \hat{\boldsymbol{\mu}}_{jk})'}{\sum_{n=1}^{N} \gamma_n(j,k)}
\end{aligned} \quad (7)$$

donde $\gamma_n(j,k)$, es la probabilidad de estar en el estado j, en el instante n con la k-ésima mezcla correspondiente a $\mathbf{x}_n$. El término $\gamma_n(j,k)$ se convierte en $\gamma_n(j)$ para el caso de una única mezcla. Para encontrar la mejor secuencia de estados, se puede emplear el algoritmo de Viterbi [9].

4. EXTENSIÓN PARA MODELOS DE PPCA EN EL TIEMPO

Una conexión natural entre la HMM y PCA proviene del hecho de que ambos, el modelo de observación para la HMM en (4) y el modelo de PPCA requieren el cálculo de una matriz de covarianza muestral dada por (3), sin embargo, debe notarse que la expresión $\hat{\mathbf{S}}$ en (7) constituye una versión pesada de (3), de manera que es preferible obtener una expresión donde los datos observados y los pesos $\gamma_n(j,k)$, pertenezcan a diferentes representaciones matriciales. Para hacer esto, partiendo de (7), se puede demostrar que [10]

$$\begin{aligned}
\hat{\mathbf{S}}_{jk} &= \sum_{n=1}^{N} r_n(j,k)(\mathbf{x}_n - \hat{\boldsymbol{\mu}}_{jk})(\mathbf{x}_n - \hat{\boldsymbol{\mu}}_{jk})' \\
&= \mathbf{X}\mathbf{R}_{jk}\mathbf{R}'_{jk}\mathbf{X}'
\end{aligned} \quad (8)$$

donde $\hat{\mathbf{S}}_{jk}$ es la matriz de covarianza muestral local para el estado j y la k-ésima componente de la mezcla. Los términos a la izquierda de (8) están definidos como

$$\begin{aligned}
r_n(j,k) &= \frac{\gamma_n(j,k)}{\sum_{n=1}^{N} \gamma_n(j,k)} \\
\hat{\boldsymbol{\mu}}_{jk} &= \sum_{n=1}^{N} r_n(j,k)\mathbf{x}_n \\
\mathbf{R}_{jk} &= (\mathbf{I} - \mathbf{r}(j,k)\mathbf{1}')\mathbf{D}_{jk}
\end{aligned}$$

donde el vector $\mathbf{r}(j,k) = [r_1(j,k), ..., r_N(j,k)]$ y la matriz $\mathbf{D}_{jk}$ es diagonal con elementos dados por $r_1(j,k)^{1/2}, ..., r_N(j,k)^{1/2}$. La matriz $\mathbf{R}_{jk}$ puede ser vista como la matriz de responsabilidad para la k-ésima componente de la mezcla en el estado j.

Debido a que $\mathbf{X}\mathbf{R}_{jk}\mathbf{R}'_{jk}\mathbf{X}' = \mathbf{R}'_{jk}\mathbf{X}'\mathbf{X}\mathbf{R}_{jk}$ en (8) y $\mathbf{X}'\mathbf{X}$ es una matriz de covarianza muestral centrada, a partir de aquí es fácil de ver que los modelos para PCA pueden ser construidos para cada componente de cada estado correspondiente en el modelo de la HMM, de modo que la estructura de los modelos de PCA contienen la información en el tiempo provista por la matriz $\mathbf{R}_{jk}$ resultante luego del proceso de entrenamiento de la HMM.

Con el fin de construir el modelo de PPCA a través del tiempo, se deben calcular evaluaciones iterativas de $\widehat{\mathbf{W}}_{jk}$ y $\hat{\sigma}^2_{jk}$ utilizando (4) y (5), pero reemplazando $\mathbf{S}$ de la ecuación (3) por $\hat{\mathbf{S}}_{jk}$ de la ecuación (8). La representación reducida hasta p dimensiones del vector observado $\mathbf{x}$ se puede calcular usando

$$\mathbf{y}_{jk} = \mathbf{M}_{jk}^{-1}\mathbf{W}'_{jk}(\mathbf{x} - \hat{\boldsymbol{\mu}}_{jk})$$

4.1. Aplicación a la Clasificación de Series de Tiempo

En este caso, los M modelos de PPCA $(\widehat{\mathbf{W}}_{jk}, \hat{\sigma}_{jk}^2)$ estimados para cada estado se emplean para transformar la secuencias de tiempo que van a ser clasificadas. Para hacer esto último, primero se utiliza el algoritmo de Viterbi para encontrar la secuencia de estados que mejor se ajusta a una serie en particular. La aplicación del algoritmo de Viterbi se hace sobre una cadena de Markov sencilla con modelos de observación consistentes en funciones de densidad de probabilidad gaussianas utilizadas en el proceso de estimación de los modelos de PPCA. Una vez se ha identificado cual estado explica mejor una observación en particular, se usa el modelo de PPCA perteneciente a ese estado para transformar la observación específica. Este proceso es aplicado a todas las secuencias involucradas en el problema de reconocimiento utilizando los modelos de la clase a la cual las series de tiempo pertenecen. En este sentido, la transformación de series de tiempo es realizada a la vez de manera supervisada y dinámica.
En esta etapa, todas las secuencias involucradas en el problema de reconocimiento han sido transformadas usando un modelo de PPCA. A continuación, estas nuevas series de tiempo pueden ser clasificadas utilizando cualquier clasificador de series de tiempo, para lo cual, de nuevo una cadena oculta de Markov puede ser empleada.

5. RESULTADOS EXPERIMENTALES

5.1 Base de datos

La base de datos fue obtenida a partir de la base de datos ST-T Europea, seleccionado secciones de las señales con eventos isquémicos y normales. Se utilizaron finalmente 48 señales con eventos isquémicos y 48 con eventos normales. Se extraen, como características de las señales, los coeficientes de mayor energía por 13 niveles de descomposición de una transformación wavelet [11]. Las wavelets que se usan corresponden a la wavelet Daubechies4 (Db4) y Symlet2 (Sym2), debido a que producen resultados aceptables con cadenas ocultas de Markov discretas sobre la misma base de datos [12].

5.2 Clasificación

Para la clasificación de las diferentes series de tiempo multivariadas (las características wavelet originales y las características wavelet transformadas usando PPCA) se usan modelos ocultos de Markov con densidades de observación gaussianas. La topología de los modelos es ergódica [9] y se examina diferente número de estados para la cadena de Markov. Se utiliza validación cruzada [13] con 4 particiones. En este trabajo, se incluyen únicamente los mejores porcentajes de clasificación, resultados detallados se pueden encontrar en [10].

5.2.1 Resultados con el espacio completo

Dos diferentes series multivariadas se usan para el entrenamiento, a saber, los coeficientes wavelet originales y los vectores wavelet transformados usando el modelo de PPCA. La dimensión del espacio dinámico transformado es igual a 13, la dimensión original de los coeficientes wavelet. Las precisiones para ambas bases de datos se muestran en las tablas 1 y 2. Las convenciones para las tablas 1 y 2 son: N_s, número de estados; **O**, la serie multivariada original; **P**, la serie multivariada original transformada usando PPCA y M, número de componentes en la mezcla de densidad de observación.

Serie	M	Precisión (%)		
		N_s =3	N_s =5	N_s =10
O	1	95.83±3.4	**96.87±4.00**	96.87±4.00
P	1	**99.90±0.10**	99.90±0.10	98.95±1.08

Tabla 1. Porcentajes de clasificación usando Sym2 con el espacio completo

La tabla 1 muestra los mejores porcentajes de clasificación de la base de datos caracterizada usando Sym2 con el espacio completo. Con el espacio transformado completo usando PPCA es posible obtener un alto porcentaje de clasificación comparado con el porcentaje de la serie sin transformar. Igualmente sucede para la base de datos caracterizada con Db4, como se observa de la tabla 2.

Serie	M	Precisión (%)		
		N_s =3	N_s =5	N_s =10
O	1	94.79±4.00	**97.92±4.17**	96.88±4.00
P	1	**99.90±0.10**	99.90±0.10	98.95±1.00

Tabla 2. Porcentajes de clasificación usando Db4 con el espacio completo

Es posible obtener una tasa de acierto del 99.90±0.10 con una complejidad menor en el modelo oculto de Markov empleado en la clasificación, 3 estados contra 5 estados en el caso sin transformación. Este resultado se confirma con ambos tipos de wavelets.

5.2.2 Resultados con el espacio reducido

Se evalúa el desempeño de clasificación de la serie multivariada transformada y reducida mediante PPCA. Se usan diferentes valores de q. Para este experimento, se usa una única componente por mezcla para el

clasificador. Las convenciones son las mismas que las de las tablas 1 y 2.
La tabla 3 muestra los mejores porcentajes de clasificación de la base de datos caracterizada usando Sym2 con el espacio reducido. El desempeño de PPCA alcanza altos porcentajes y una estabilidad aceptable (en términos de la desviación estándar) al ir aumentando el número de componentes, obteniendo con 11 componentes porcentajes comparables con los de las series sin transformar.

q	N_s	Precisión (%)
3	3	61.45±10.95
5	5	84.37±5.24
7	5	**91.66±5.9**
9	10	89.58±5.37
11	10	**94.79±2.08**

Tabla 3. Porcentajes de clasificación usando Sym2 con el espacio reducido.

Resultados similares se obtienen para la base de datos caracterizada con Db4, como se observa de la tabla. Sin embargo, los porcentajes de acierto usando la serie caracterizada con Db4, son en promedio mayores que los obtenidos con Sym2, como un examen detallado de las tablas 3 y 4 lo muestra.

q	N_s	Precisión (%)
3	5	63.54±25.08
5	5	81.25±15.40
7	3	**94.79±5.24**
9	10	**97.91±4.16**
11	5	**98.95±2.08**

Tabla 4. Porcentajes de clasificación usando Db4 con el espacio reducido.

Para ambas caracterizaciones, un factor de reducción de 8/13 (61%) se obtiene si se establece como porcentaje de clasificación aceptable, un porcentaje mayor al 80%. Si el porcentaje mínimo que se tolera es del 90%, es posible obtenerlo con un factor de reducción de 7/13 (46%).

6. DISCUSIÓN

En esta trabajo se ha mostrado cómo un modelo tradicional de reducción de dimensionalidad estático, el análisis probabilístico de componentes principales, puede incorporarse a un modelo de análisis de variabilidad en el tiempo, para realizar clasificación y reducción de características dinámicas.

La capacidad de representación del algoritmo se mostró sobre una base de datos de patologías de ECG, para la cual se obtuvieron altos porcentajes de precisión incluso con dimensión reducida.

Por el lado de la reducción de características, el análisis de PCA se ha usado tradicionalmente de forma estática como parte del preproceso, por ejemplo en tareas de reconocimiento de voz [14,15], sin tener en cuenta la variabilidad de las características a través del tiempo. El modelo propuesto en este trabajo generaliza el enfoque anterior en el caso en que el modelo de reducción markoviano no tiene un único estado.

Aunque la metodología de reducción y clasificación propuesta en este trabajo ha permitido obtener porcentajes considerables en la reducción y altas tasas de acierto en la clasificación, algunas preguntas deben resolverse primero a fin de hacer más sólido el enfoque. Tal vez la más difícil de todas tiene que ver con la selección del modelo. Primero, el número de estados en la cadena de Markov que mejor explique el comportamiento dinámico debe ser escogido cuidadosamente, ya que del número de estados depende el número de parámetros de los modelos de PCA que deben estimarse. Segundo, debe encontrarse una manera óptima de encontrar la dimensión del espacio latente.

7. AGRADECIMIENTOS

Este trabajo se realiza en el marco del proyecto *Sistema automatizado de clasificación de eventos fisiológicos a partir de patrones bioeléctricos como soporte en el tratamiento de la enfermedad de Parkinson y otros desórdenes neurológicos* financiado por Colciencias, código 1101417904 y bajo el marco del proyecto *Selección de características dinámicas empleando procesos markovianos, aplicadas al reconocimiento de disfunciones en bioseñales,* financiado por la convocatoria para la financiación de tesis de maestría año 2005-2006 del Centro de Investigaciones y Extensión de la Universidad Tecnológica de Pereira.

BIBLIOGRAFÍA

[1] I.T. Jolliffe, Principal Component Analysis, Springer Verlag, second edition, 2002.

[2] M.E. Tipping y C.M. Bishop, Probabilistic Principal Component Analysis, Journal of the Royal Statistical Society, Series B, vol. 21, no. 3, pp. 611–622, 1999.

[3] M.A. Carreira-Perpiñán, Continuous Latent Variable Models for Dimensionality Reduction and Sequential Data Reconstruction, Ph.D. thesis, Department of Computer Science, University of Sheffield, UK, 2001.

[4] M.E. Tipping y C.M. Bishop, Mixture of Probabilistic Principal Component Analysers, Neural Computation, vol. 11, no. 2, pp. 443–482, 1999.

[5] Alpert JS y et. al., Myocardial Infarction Redefined -- A Consensus Document of The Joint European Society of Cardiology/American College of Cardiology Committee for the Redefinition of Myocardial Infarction. J Am Coll Cardiol. 2000 Sep;36(3):959-69.

[6] Bedoya, T y et. al., Estudio Descriptivo sobre Infarto Agudo de Miocardio en el Hospital de Caldas ESE entre 1996-2002. Revista Colombia Médica. Vol 35 No 3. Año 2004. Universidad del Valle.

[7] W. Härdle y L. Simar, Applied Multivariate Statistical Analysis, Springer, N.Y., 2003.

[8] X. Huang, A. Acero, y H.W. Hon, Spoken Language Processing: A Guide to Theory, Algorithm and System Development, Prentice Hall, Upper Saddle River. New Jersey, first edition, 2001.

[9] L. R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of The IEEE, vol. 77, no. 2, 1989.

[10] Alvarez M. Reducción de Dimensión de características Dinámicas empleando Procesos Markovianos Aplicados al Reconocimiento de Disfunciones en Bioseñales, Tesis de Maestría, Universidad Tecnológica de Pereira, Pereira, Colombia. 2006.

[11] Mallat S. A Wavelet Tour of Signal Processing. San Diego, California: Academic Press, 1998.

[12] Arias J.D., Alvarez M., Castellanos G. y Godino J.I., Caracterización Dinámica de Señales de ECG con Infarto Agudo de Miocardio usando HMM, en Libro de Actas del XXIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica, Noviembre 2005, pp. 431--433.

[13] Theodoridis S. y Koutroumbas, K. Pattern Recognition. Academic Press, 1999.

[14] Jankowski C. R., Do H.-D. H. y Lippmann , R., A Comparison of Signal Processing Front-Ends for Automatic Word Recognition, IEEE Transactions on Speech and Audio Processing, vol. 3, no.4, Julio 1995.

[15] Alvarez, M., Castellanos, G., Suárez J.F., y Godino, J. I. Identificación de Patologías de Voz usando HMM, en Libro de Actas del XXII Congreso Anual de la Sociedad Española de Ingeniería Biomédica, Noviembre, 2004, pp. 217--220.