

ANÁLISIS DE PROXIMIDAD DE MODELOS OCULTOS DE MARKOV PARA LA IDENTIFICACIÓN DE FUENTES DE ESPIGAS

RESUMEN

Recientemente, los modelos ocultos de Markov (HMM) se han usado para identificar fuentes de espigas en el tratamiento de la enfermedad de Parkinson. El criterio de clasificación que suele emplearse es la regla MAP (máximo a posteriori) para reconocer la clase correcta. Sin embargo, la clasificación puede mejorarse usando análisis de proximidad transformando las matrices que caracterizan el proceso de Markov a un espacio en el que se ven mejor reflejadas las similitudes y las diferencias entre los parámetros de estas matrices. Se presenta la aplicación del enfoque de análisis de proximidad usando HMM en la identificación de fuentes de espigas (Tálamo y Subtálamo, GPi y GPe). Los resultados muestran que el análisis de proximidad entrega mejores resultados (4% en promedio) que el enfoque tradicional.

PALABRAS CLAVES: Enfermedad de Parkinson, Modelos ocultos de Markov, Análisis de Proximidad, Análisis de Componentes Principales.

ABSTRACT

Recently, Hidden Markov Models (HMM) have been used for the identification of spike sources in Parkinson's disease treatment. Usual classification criteria consists in maximum a posteriori rule (MAP) for the recognition of the correct class. However, classification could be improved using proximity analysis, transforming matrices of Markov process to another space where similarities and differences are better revealed. The authors present the proximity analysis approach using HMM for the identification of spike sources (Thalamo and Subthalamo, Gpi and GPe). Results show that proximity analysis improves recognition performance for about 4% over traditional approach.

KEYWORDS: Parkinson's Disease, Hidden Markov Models, Proximity Analysis, Principal Component Analysis.

1. INTRODUCCIÓN

Con el fin de reducir los síntomas inherentes a la enfermedad de Parkinson (rigidez, tremor) suele recurrirse a un procedimiento quirúrgico que consiste en la implantación de micro-estimuladores en el STN (núcleo subtalámico), GPi (glóbulo pálido interno) o el Vim (núcleo ventro-intermedio del tálamo) [1].

Una de las tareas más difíciles en este tipo de cirugía consiste en identificar la región del cerebro en la cual se debe ubicar el micro estimulador, tarea que tradicionalmente se ha resuelto usando el criterio subjetivo del neurocirujano que, con base en la acústica de las señales provenientes de las diferentes zonas, decide en qué punto se encuentra el electrodo [2]. Aunque este diagnóstico suele realizarse además con la ayuda de tomografías y un software de planeamiento, la localización exacta del micro-electrodo sigue siendo bastante imprecisa.

El establecimiento de un sistema automatizado que evalúe de manera objetiva la ubicación de la sonda en el cerebro permanece como un campo abierto de trabajo debido a la dificultad que representa la caracterización

MAURICIO ÁLVAREZ

Ingeniero Electrónico, MsC.
Profesor
Universidad Tecnológica de Pereira
malvarez@ohm.utp.edu.co

GERMÁN CASTELLANOS

Ph.D. en Telecomunicaciones
Docente de planta
Universidad Nacional
de Colombia, s. Manizales
gcastell@telesat.com.co

ÁLVARO OROZCO

MsC. en Ingeniería Eléctrica
PhD (c) en Bioelectrónica
Docente Titular
Universidad Tecnológica de Pereira
aaog@utp.edu.co

adecuada de los trenes de espigas. En [3], los autores proponen el uso de modelos ocultos de Markov para caracterizar la dinámica de cambio de los trenes de espigas, obteniendo porcentajes de acierto de clasificación superiores a otros métodos propuestos con anterioridad en la literatura. En los sistemas de clasificación basados en HMM, la identificación de las clases se suele realizar caracterizando las señales como una secuencia de vectores de características y entrenando por cada clase un modelo oculto de Markov. Para una nueva observación, por cada HMM se calcula la verosimilitud y el modelo que entregue la verosimilitud mayor se escoge como el modelo que representa la clase correcta.

En este artículo se presenta un método de clasificación alternativo que utiliza análisis de proximidad para hacer la clasificación de los modelos ocultos de Markov y que ha sido usado con éxito en la clasificación de otro tipo de bioseñales [4]. La clasificación basada en similitudes o análisis de proximidad difiere de los enfoques típicos de reconocimiento de patrones donde los objetos son representados por conjuntos (vectores) de características. En la clasificación basada en similitudes, los objetos se describen usando similitudes o disimilitudes por

pares, es decir, distancias a otros objetos en el conjunto de datos.

La idea en la base del enfoque propuesto es conceptualmente simple: construir un nuevo espacio de representación, usando los valores de similaridad entre secuencias obtenidos mediante HMMs y diseñar un clasificador en tal espacio. Una de las justificaciones de este enfoque se encuentra en el hecho de que la similaridad es alta para secuencias que comparten características parecidas, es decir, que pertenecen a la misma clase, y baja para objetos de diferentes clases, haciendo posible la discriminación.

El problema principal en este enfoque se encuentra en la alta dimensionalidad que suelen tener los vectores en el espacio de similaridad. Una forma directa de solucionar este problema consiste en usar análisis multivariado [5] para reducir el conjunto inicial de entrenamiento.

En la sección 2 se describe el enfoque de análisis de proximidad. En la sección 3 se hace un corto resumen de los modelos ocultos de Markov. En la sección 4 se describe la transformación de las matrices estocásticas del HMM al espacio de similaridad. En la sección 5 se presentan los resultados de aplicar el enfoque descrito para reconocer fuentes de espigas en dos bases de datos distintas, a saber: entre tálamo y subtálamo, y entre Border, Burster, GPi, Pauser y Striatum.

2. ANÁLISIS DE PROXIMIDAD

La clasificación basada en similaridades o análisis de proximidad difiere de los enfoques típicos de reconocimiento de patrones donde los objetos son representados por conjuntos (vectores) de características. En la clasificación basada en similaridades, los objetos se describen usando similaridades o diferencias por pares, es decir, distancias a otros objetos en el conjunto de datos.

Un valor de similaridad expresa la magnitud de diferencia entre dos objetos y se vuelve cero sólo cuando los objetos son idénticos.

Una forma directa de trabajar con las representaciones de disimilaridad está basada en las relaciones de distancia entre los objetos, que naturalmente conduce a métodos basados en rango.

Los datos de distancia también pueden tratarse como la descripción de un espacio característico específico donde cada dimensión corresponde a un objeto. Lo anterior no cambia necesariamente el enfoque tradicional con base en características, aunque envuelve el problema en un caso particular cuando el número de muestras N se hace igual a su dimensionalidad R . En general cualquier clasificador que opere en el espacio de características puede usarse. En el proceso de aprendizaje, las reglas de

decisión se construyen sobre una matriz $N \times N$ de distancias, por pares de puntos.

El problema principal del enfoque basado en similaridades, de particular relevancia en aplicaciones prácticas, es la alta dimensionalidad del espacio de similaridad resultante. Se han propuesto dos tipos de soluciones para resolver este problema. El primero consiste en construir el espacio de similaridad usando todos los patrones disponibles, y a continuación aplicar algunas técnicas estándar de reducción de dimensionalidad como por ejemplo, escalamiento multidimensional [6]. El segundo tipo de solución escoge un conjunto de prototipos de cada clase utilizando algún criterio de selección y la clasificación se realiza con respecto a estos prototipos [7].

3. MODELOS OCULTOS DE MARKOV

Un HMM es básicamente una cadena de Markov donde la observación de salida es una variable aleatoria generada de acuerdo a una función de probabilidad asociada a cada estado. Formalmente hablando, un modelo oculto de Markov está definido por la matriz \mathbf{A} , la matriz de transición de probabilidades, la matriz \mathbf{B} , matriz de probabilidad de salida y el vector $\boldsymbol{\pi}$, que representa la distribución inicial de los estados.

Por conveniencia se usa la siguiente notación

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

para indicar el conjunto total de parámetros de un HMM.

El enfoque típico para realizar la clasificación cuando se usan HMM adopta el criterio de máxima verosimilitud, donde una secuencia desconocida \mathbf{O} se le asigna a la clase que muestre la verosimilitud mayor, es decir,

$$Clase(\mathbf{O}) = \arg \max_k P(\mathbf{O} | \Phi_k)$$

donde Φ_k es el HMM correspondiente a la k -ésima clase. Lo anterior requiere el entrenamiento de C HMMs para un problema de C clases. En lo que sigue, a este enfoque se le designará como el enfoque ML_{UPC} (donde UPC significa "uno por clase").

Otro regla de clasificación diferente se describe a continuación. En lugar de entrenar un HMM por cada clase, se entrena un modelo por cada secuencia de entrenamiento y se asigna una secuencia desconocida \mathbf{O} a la clase del modelo que muestra la más alta verosimilitud. Formalmente, sea $\Phi_k^{(p)}$ el modelo entrenado con la secuencia $\mathbf{O}_k^{(p)}$, que pertenece a la clase p . La regla de clasificación bajo este enfoque es:

$$Clase(\mathbf{O}) = \arg \max_p \left(\max_k P(\mathbf{O} | \Phi_k^{(p)}) \right)$$

A este enfoque se le llamará el enfoque ML_{UPS} (donde UPS significa “uno por secuencia”).

4. APLICACIÓN A LOS MODELOS OCULTOS DE MARKOV

La cuestión básica de la estrategia basada en similaridad está en cómo definir las en el contexto de un HMM. Como es sabido, dado un HMM Φ y una secuencia \mathbf{O} , existe un método estándar (el procedimiento de forward-backward) para calcular $P(\mathbf{O}|\Phi)$, es decir, la probabilidad de que la secuencia fue generada por el modelo Φ . Esta medida cuantifica la llamada verosimilitud y mide qué tan bien la secuencia \mathbf{O} se acomoda al modelo Φ . Una opción natural consistiría entonces en definir la similaridad $D_{ij} = \Delta(\mathbf{O}_i, \mathbf{O}_j)$ entre dos secuencias \mathbf{O}_i y \mathbf{O}_j como [8,9]

$$D_{ij} = \Delta(\mathbf{O}_i, \mathbf{O}_j) = \frac{\log P(\mathbf{O}_i | \Phi_j)}{T_i}$$

donde Φ_j es el HMM entrenado con la secuencia \mathbf{O}_j y T_i es la longitud de la secuencia \mathbf{O}_i . El $1/T_i$ es el factor de normalización introducido para tener en cuenta secuencias de diferente longitud. Nótese que estas similaridades no son simétricas.

4.1 Definición formal

Considérese un problema de clasificación con C clases. Para cada clase $k \in \{1, 2, \dots, C\}$ se tiene un conjunto de N_k secuencias de entrenamiento $\mathfrak{S}_k = \{\mathbf{O}_1^{(k)}, \mathbf{O}_2^{(k)}, \dots, \mathbf{O}_{N_k}^{(k)}\}$; así, $N = \sum_k N_k$ es el tamaño total del conjunto de entrenamiento $\mathfrak{S} = \sum_k \mathfrak{S}_k$.

Sea $\mathfrak{R} = \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_R\}$ un conjunto R de referencias u objetos representativos; estos objetos pueden pertenecer a un conjunto de secuencias de entrenamiento ($\mathfrak{R} \in \mathfrak{S}$) o puede definirse de otra forma. Ahora, sea $\Delta_{\mathfrak{R}}(\mathbf{O})$ una función que retorna el vector de similaridades entre una secuencia arbitraria \mathbf{O} y todas las secuencias en \mathfrak{R} , así:

$$\Delta_{\mathfrak{R}}(\mathbf{O}) = \begin{bmatrix} \Delta(\mathbf{O}, \mathbf{P}_1) \\ \vdots \\ \Delta(\mathbf{O}, \mathbf{P}_R) \end{bmatrix} \in \mathbb{R}^{\mathfrak{R}}$$

Se designará el espacio $\mathbb{R}^{\mathfrak{R}}$ en el cual el vector de similaridad existe como el “espacio de similaridad” y se denotará por $\mathfrak{S}_{\mathfrak{R}}$, donde el subíndice \mathfrak{R} se usa para enfatizar la dependencia del espacio de similaridad con respecto al conjunto \mathfrak{R} . Una vez definido este espacio de similaridad, en principio, se puede usar cualquier clasificador estándar.

Dependiendo del valor de \mathfrak{R} , se pueden tener diferentes enfoques; el básico consiste en escoger $\mathfrak{R} = \mathfrak{S}$, el conjunto total de entrenamiento. Con esta opción la dimensionalidad de $\mathfrak{S}_{\mathfrak{R}} = \mathfrak{S}_{\mathfrak{S}}$ es igual a N , la dimensionalidad del conjunto de entrenamiento \mathfrak{S} .

5. MARCO EXPERIMENTAL

5.1 Bases de Datos

Las bases de datos utilizadas en los experimentos se muestran en la tabla 1. La base de datos **DB1** incluye muestras de las zonas Tálamo y Subtálamo, en la trayectoria hacia el Subtálamo. La base de datos **DB2** incluye muestras de la zona GPi y de varias zonas Border, Burster, Pauser y Striatum en la trayectoria hacia el GPi y conocidas de forma general como GPe (glóbulo pálido externo).

Base de Datos	Zona Cerebral	Muestras
DB1	Tálamo	50
	Subtálamo	50
DB2	Border	8
	Burster	8
	GPi	10
	Pauser	5
	Striatum	3

Tabla 1. Bases de datos de zonas cerebrales usadas en los experimentos.

5.2 Extracción de Características

Las características extraídas se obtienen usando la *transformada wavelet* [10]. Cada señal es segmentada en ventanas de longitud igual a 250 ms con un traslape de 50 ms. Para cada marco se aplica la transformada wavelet Haar y por cada escala de detalle se selecciona el coeficiente con la amplitud mayor para ser utilizado como parte del vector de observación. El procedimiento de caracterización se realiza de este modo con el fin de aprovechar la propiedad de compactación de energía que poseen las wavelets.

5.2 Topología de los HMM

Para disminuir la complejidad computacional, se utilizan modelos ocultos de Markov discretos para representar diferentes zonas cerebrales. Se utilizan modelos de topología ergódica con cinco y diez estados y 64 y 128 palabras para el libro de códigos. El grupo de parámetros dado en

¡Error! No se encuentra el origen de la referencia., y que describe el modelo HMM de cada tipo de zona se estima utilizando el algoritmo estándar de Máximación de la Esperanza (*EM Expectation-Maximization*) [11].

5.3 Reducción de dimensionalidad

Como se dijo anteriormente, en el enfoque ML_{UPS} , los datos de entrenamiento en el espacio de similaridad tienen una alta dimensionalidad. En este trabajo se ha usado únicamente PCA para hacer la reducción de dimensión, debido a que el objetivo del trabajo no es tanto los esquemas de reducción sino mostrar los resultados en el espacio de similaridad.

5.4 Validación

Se comparan los resultados de usar el enfoque ML_{UPC} y el enfoque ML_{UPS} usando dos clasificadores paramétricos: un clasificador lineal basado en una densidad normal regularizada (RLNC) y un clasificador cuadrático basado en una densidad normal regularizada (RQNC); y tres clasificadores no paramétricos: un clasificador de Parzen optimizado [12] (PARZEN), un clasificador de vecino más cercano (1-NN) y un clasificador de 3 vecinos más cercanos (3-NN) [13]. Se utiliza validación leave-one-out y se repite el experimento 5 veces.

ENFOQUE	5 ESTADOS		10 ESTADOS	
	$\mu(\%)$	$\sigma(\%)$	$\mu(\%)$	$\sigma(\%)$
ML_{UPC}	95.8	1.79	96	3.08
RLNC	99.4	0.5	99.9	0.01
RQNC	99.8	0.5	99	0
PARZEN	99.9	0.01	99.9	0.01
1-NN	99.9	0.01	99.9	0.01
3-NN	97.4	0.5	97.6	0.58

Tabla 2. Tasa de acierto para la base de datos **DB1** usando el enfoque ML_{UPC} y el enfoque ML_{UPS} para diferente número de estados y 64 palabras en el libro de códigos.

5.5 Resultados

La tabla 2 muestra la tasa de acierto obtenida mediante los diferentes enfoques de clasificación usando la base de datos **DB1** y 64 palabras en el libro de códigos.

ENFOQUE	5 ESTADOS		10 ESTADOS	
	$\mu(\%)$	$\sigma(\%)$	$\mu(\%)$	$\sigma(\%)$
ML_{UPC}	97.2	1.3	97.4	2.19
RLNC	99	0	99	0
RQNC	99	0	99.4	0.96
PARZEN	99.9	0.01	99.9	0.01
1-NN	99.9	0.01	99.9	0.01
3-NN	97	0	0.98	0

Tabla 3. Tasa de acierto para la base de datos **DB1** usando el enfoque ML_{UPC} y el enfoque ML_{UPS} para diferente número de estados y 128 palabras en el libro de códigos.

La tabla 3 muestra la tasa de acierto obtenida mediante los diferentes enfoques de clasificación usando la base de datos **DB1** y 128 palabras en el libro de códigos.

De las tablas 2 y 3 se observa que, independientemente del número de palabras que se usen en el libro de códigos y del número de estados, los clasificadores no paramétricos, en particular, el clasificador optimizado de Parzen y el clasificador de un vecino más cercano, permiten obtener los mejores porcentajes de acierto en la clasificación de las zonas tálamo y subtálamo.

La figura 1 muestra la media de la tasa de acierto de la clasificación obtenida al aplicar PCA al conjunto de muestras en el espacio de similaridad. Para este experimento se usó la topología de 5 estados y 64 palabras en el libro de códigos. En la tabla 4 se resumen los mejores porcentajes entregados por cada clasificador.

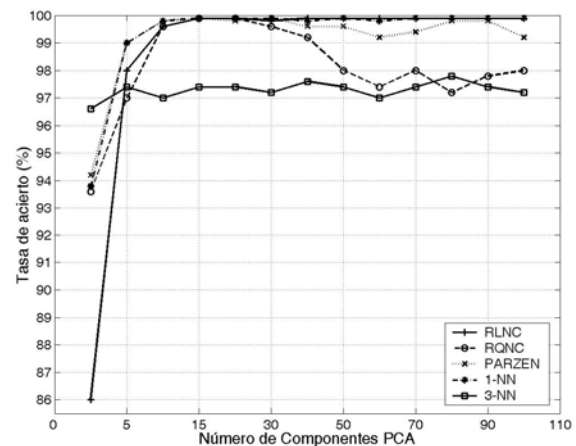


Figura 1. Tasa de acierto en promedio contra el número de componentes en PCA para la base de datos **DB1**.

Clasificador	Componentes	$\mu(\%)$	$\sigma(\%)$
RLNC	15	99.9	0.01
RQNC	15	99.9	0.01
PARZEN	15	99.9	0.01
1-NN	15	99.9	0.01
3-NN	80	97.8	0.5

Tabla 4. Número de componentes y mejores porcentajes de clasificación usando PCA para la base de datos **DB1**.

De la tabla 4 se observa que la mayoría de clasificadores permiten obtener un porcentaje muy alto de clasificación (99.9%) con tan sólo 15 componentes. Por debajo de 15 componentes, el clasificador de Parzen y el clasificador de 1-NN permiten obtener porcentajes de acierto de 99% con 5 componentes. De la figura 1, se observa que con sólo un componente PCA es posible obtener porcentajes de clasificación por encima del 90% (no se cumple en el caso del clasificador RLNC).

La tabla 5 muestra la tasa de acierto obtenida mediante los diferentes enfoques de clasificación usando la base de datos **DB2** y 64 palabras en el libro de códigos. La tabla 6 muestra la tasa de acierto obtenida mediante los diferentes enfoques de clasificación usando la base de datos **DB2** y 128 palabras en el libro de códigos.

ENFOQUE	5 ESTADOS		10 ESTADOS	
	μ (%)	σ (%)	μ (%)	σ (%)
ML_{UPC}	94.12	3.6	89.41	5.34
RLNC	92.94	3.35	93.53	3.22
RQNC	87.06	2.63	83.53	3.35
PARZEN	97.06	2.08	97.06	2.08
1-NN	96.47	3.22	95.88	3.35
3-NN	85.29	2.94	82.94	2.46

Tabla 5. Tasa de acierto para la base de datos **DB2** usando el enfoque ML_{UPC} y el enfoque ML_{UPS} para diferente número de estados y 64 palabras en el libro de códigos.

Igual que en el caso de la base de datos **DB1**, las tablas 5 y 6 permiten observar que en la base de datos **DB2** los porcentajes de clasificación mayores se obtienen usando los clasificadores no paramétricos optimizado de Parzen y vecino más cercano. Las desviaciones estándar en ambos casos son menores que para los demás clasificadores.

ENFOQUE	5 ESTADOS		10 ESTADOS	
	μ (%)	σ (%)	μ (%)	σ (%)
ML_{UPC}	95.88	2.63	92.94	4.46
RLNC	94.71	1.32	91.76	3.83
RQNC	87.65	1.32	83.53	4.46
PARZEN	97.06	2.08	94.71	3.83
1-NN	96.47	2.46	95.29	3.35
3-NN	95.29	2.63	92.35	3.9

Tabla 6. Tasa de acierto para la base de datos **DB2** usando el enfoque ML_{UPC} y el enfoque ML_{UPS} para diferente número de estados y 128 palabras en el libro de códigos.

La figura 2 muestra la media de la tasa de acierto de la clasificación obtenida al aplicar PCA al conjunto de muestras en el espacio de similaridad. Para este experimento se usó la topología de 5 estados y 128 palabras en el libro de códigos. En la tabla 7 se resumen los mejores porcentajes entregados por cada clasificador.

La tabla 7 permite observar que el mejor clasificador se obtiene utilizando el clasificador no paramétrico de un vecino más cercano. Para este caso, con sólo 3 componentes es posible lograr el más alto porcentaje de acierto, igual al 98.24%. El clasificador de Parzen permite obtener, igualmente, un porcentaje alto de clasificación correcta.

4. CONCLUSIONES

En general los resultados de acierto en la clasificación usando análisis de proximidad son mayores que los obtenidos usando el enfoque de máxima verosimilitud cuando las señales provenientes de diferentes zonas cerebrales se caracterizan usando modelos ocultos de Markov.

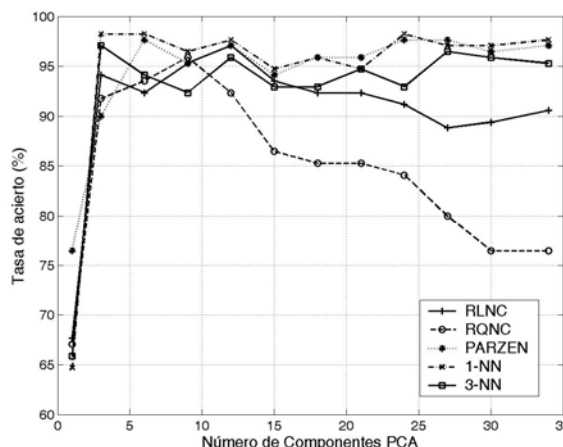


Figura 2. Tasa de acierto en promedio contra el número de componentes en PCA para la base de datos **DB2**.

Clasificador	Componentes	μ (%)	σ (%)
RLNC	12	97.06	2.08
RQNC	9	95.88	1.61
PARZEN	6	97.65	2.46
1-NN	3	98.24	2.63
3-NN	3	97.06	2.08

Tabla 7. Número de componentes y mejores porcentajes de clasificación usando PCA para la base de datos **DB2**.

En el espacio de similaridad, los clasificadores no paramétricos permiten obtener porcentajes de clasificación mejores. En particular, para la base de datos **DB1**, el clasificador de vecino más cercano y el optimizado de Parzen alcanzan los porcentajes más altos de clasificación junto con un factor de reducción menor (15/100), mientras que en la base de datos **DB2**, el clasificador por vecino más cercano tiene un mejor rendimiento con un factor de reducción de 3/34.

Se concluye igualmente que usando PCA se logra reducir adecuadamente el número de componentes necesarios para lograr altos porcentajes de clasificación.

Como trabajo futuro se propone aumentar el número de muestras en la base de datos **DB2** y usar otras distancias entre HMMs junto con otros esquemas diferentes de reducción de dimensionalidad.

5. 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.

6. BIBLIOGRAFÍA

- [1] TEXEIRA, M. F, FONOFF, E. “Neurosurgical treatment of movement disorders.” *Rev Med (Sao Paulo)*, vol. 83, pp. 1–16, 2004.
- [2] FALKENBERG, J. H. McNAMES, J., BURCHIEL, K. J. “Statistical methods of analysis and visualization of extracellular microelectrode recordings,” *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2003, pp. 2515-2518.
- [3] OROZCO, A.A., GUARNIZO, C. ÁLVAREZ, M., CASTELLANOS, G. GUIJARRO, E. “Spike Source Identification using Artificial Intelligence Techniques”. *Proceedings of The 3RD IEE International Seminar on Medical Applications of Signal Processing*. Londres. Noviembre de 2005. págs. 105-109.
- [4] ÁLVAREZ, M., ARIAS, J.D., CASTELLANOS, G., GODINO, J.I. “Identificación de voces disfuncionales usando HMM y análisis de proximidad”. *Memorias del XXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica, CASEIB 2005*. Madrid. Noviembre. 2005. págs. 125-128
- [5] JOHNSON, R. A. y WICHERN, D. W., *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, NJ.07458, 2002.
- [6] PEKALSKA, E. DUIN, R. “Spatial representation of dissimilarity data via lower-complexity linear and non linear mappings,” *T. Caelli, A. Amin, R.P.W. Duin, M. Kamel, D. de Ridder (Eds.), Structural, Syntactic and Statistical Pattern Recognition, Lecture Notes in Computer Science*, vol. 2396, pp. 488–497, 2002.
- [7] PEKALSKA, E. DUIN, R. “Dissimilarity representations allow for building good classifiers”, *Pattern Recognition Letters*, vol. 23, no. 8, pp. 943–956, 2002.
- [8] BICEGO, M., MURINO, V., FIGUEIREDO, M., “Similarity-based clustering of sequences using hidden Markov models,” *Proceedings of the International Conference on Machine Learning and Data Mining*, 2003.
- [9] BICEGO, M., MURINO, V., FIGUEIREDO, M., “Similarity-based classification of sequences using hidden markov models,” *Pattern Recognition*, vol. 37, Diciembre 2004.
- [10] MALLAT, S. *A Wavelet Tour of Signal Processing*. Upper Saddle River, New Jersey: Prentice Hall, 2001.
- [11] McLACHLAN, G.J., KRISHNAN, T. *The EM Algorithm and its Extensions*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. New York. 1997.
- [12] VAN DER HEIJDEN, F., DUIN, R., De RIDDER, D., TAX, D.M.J, *Classification, parameter estimation and state estimation-an engineering approach using Matlab*. John Wiley & Sons. 2004.
- [13] DUDA, R. O., HART, P. E., y STORK, D. G., *Pattern Classification*, Segunda ed. John Wiley & Sons, INC, 2001.