

MODELOS OCULTOS DE MARKOV EN ESPACIOS DE DISIMILARIDAD: ALTERNATIVAS PARA LA SELECCIÓN DE PROTOTIPOS

Hidden Markov models in dissimilarity spaces: alternatives for prototype selection.

RESUMEN

El criterio convencional de clasificación en sistemas que involucran modelos ocultos de Markov emplea la regla de máxima verosimilitud para escoger la clase correcta. Existe evidencia que muestra que la clasificación basada en disimilaridades entre modelos ocultos de Markov aumenta el desempeño del sistema. En este nuevo espacio de disimilaridades, las reglas de decisión pueden construirse usando todo el conjunto de entrenamiento o un conjunto reducido de prototipos adecuadamente seleccionados, que permiten minimizar el número de disimilaridades que deben medirse. En este artículo se comparan diferentes procedimientos para la selección de prototipos en el espacio de disimilaridades entre modelos ocultos de Markov.

PALABRAS CLAVES: espacio de disimilaridad, modelos ocultos de Markov, selección de prototipos.

ABSTRACT

Traditional classification criteria in systems involving hidden Markov models employs maximum likelihood rule for the election of the correct class. There is evidence showing that a dissimilarity-based classification scheme between hidden Markov models improves the system's performance. In this new space, decision rules can be built using the whole training set or a reduced number of prototypes carefully chosen, so that just a few dissimilarities must be measured. In this paper, we compare different procedures for prototype selection in a dissimilarity space designed using hidden Markov models.

KEYWORDS: *dissimilarity space, hidden Markov model, prototype selection.*

1. INTRODUCCIÓN

Los modelos ocultos de Markov (hidden Markov models - HMM) han sido empleados con éxito en el modelado de secuencias temporales. En particular, se han utilizado en diversas tareas de reconocimiento de voz [1] y en los últimos años su uso se ha extendido a aplicaciones que involucran señales biomédicas [2]. Cuando se emplean HMM para hacer reconocimiento de patrones, el enfoque estándar de clasificación consiste en entrenar un HMM por cada clase y luego emplear los HMM como funciones de probabilidad de clase condicional en el paradigma clásico de clasificación bayesiano. Si se asume que las clases tienen probabilidades a priori iguales, la categoría de una secuencia desconocida se determina como aquella cuyo modelo entrega la probabilidad más alta de haber generado esa secuencia. Esta regla de clasificación se conoce como la regla de máxima verosimilitud (maximum likelihood - ML) [1].

De otro lado, la forma convencional de representar los objetos en un problema de reconocimiento de patrones es mediante el empleo de vectores de características. Sin embargo, en algunos casos obtener una descripción apropiada de los objetos en términos de características es

difícil o, si se obtiene, esta descripción puede ser ineficiente para el aprendizaje de los parámetros del clasificador. Una representación alternativa se basa en las relaciones de similaridad o disimilaridad entre objetos. Cuando esta representación se define apropiadamente, puede ser ventajosa para resolver problemas de identificación de clases [3].

Para los sistemas de reconocimiento que involucran HMM es posible realizar la clasificación usando un enfoque alternativo a la regla de máxima verosimilitud mediante el empleo de espacios de disimilaridad. Usando distancias entre pares de HMM, la matriz de disimilaridades se construye como un conjunto organizado de estos valores de distancia y la clasificación se realiza sobre esta matriz de disimilaridad. Esta forma alternativa de clasificación ha mostrado mejores porcentajes de acierto que la clasificación tradicional usando la regla ML, para aplicaciones de identificación de patologías de voz [4] y de reconocimiento de formas en objetos de dos dimensiones [5].

En el espacio de disimilaridades, el clasificador puede entrenarse usando el espacio completo de las observaciones (vectores de distancias) o sólo un conjunto reducido seleccionado adecuadamente. La ventaja

MAURICIO ÁLVAREZ

Ingeniero Electrónico, M.Sc.

Profesor Auxiliar

Universidad Tecnológica de Pereira

malvarez@utp.edu.co

RICARDO HENAO

Ingeniero Electrónico, M.Sc.

Profesor Auxiliar

Universidad Tecnológica de Pereira

rhenao@utp.edu.co

principal de la selección de prototipos consiste en minimizar el número de disimilaridades que deben medirse para la clasificación de nuevas señales, representadas por los parámetros del HMM. Adicionalmente, la selección de prototipos permite una reducción de dimensionalidad del espacio de entrenamiento, evitando de esta forma el conocido problema de la maldición de la dimensionalidad [6].

En este artículo, se presentan diferentes alternativas para la selección de prototipos en el espacio de disimilaridad construido a partir de distancias entre los parámetros de HMM. En particular se presentan cuatro procedimientos AleatoriaC, Aleatorio, R-CentrosC y R-Centros, empleados comúnmente en el estado del arte [6].

La aplicación de la metodología propuesta se realiza sobre señales provenientes de microelectrodos de registro, obtenidas a partir de procedimientos quirúrgicos empleados para introducir microestimuladores en ciertas áreas específicas del cerebro, con el objetivo de desinhibir síntomas en pacientes con enfermedad de Parkinson.

2. MODELOS OCULTOS DE MARKOV EN EL ESPACIO DE DISIMILARIDAD

2.1 Análisis de proximidad

El primer paso en un problema de reconocimiento de patrones es la búsqueda de una representación apropiada de los objetos a clasificar. Luego, es posible construir una regla de decisión, que discrimine entre las diferentes categorías y proporcione una buena capacidad de generalización. Convencionalmente, los objetos se representan mediante el empleo de rasgos. En algunos casos, sin embargo, es difícil encontrar una representación apropiada de los objetos en términos de sus características. Una representación alternativa se basa en las relaciones de disimilaridad entre objetos. Cuando se define de manera apropiada, esta representación puede tener ventajas para resolver problemas de clasificación [3].

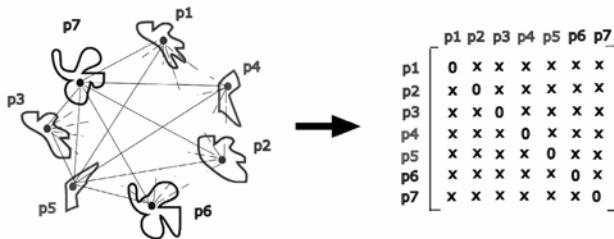


Figura 1. Construcción de la matriz de disimilaridades a partir de la medición de distancias entre diferentes objetos. La distancia por pares de objetos se introduce en la matriz. La distancia de un objeto a sí mismo es cero.

La idea general detrás del análisis de proximidad consiste en que dado un conjunto de valores de disimilaridades entre objetos, es posible construir un nuevo espacio de representación en el que cada objeto se describe por estos

valores. La clasificación se realiza entonces en este nuevo espacio. Varios estudios experimentales han mostrado que la filosofía de análisis de proximidad mejora el desempeño de clasificación en tareas de alta complejidad [3].

La figura 1 muestra la forma de construir el espacio de disimilaridad mediante la medición de distancias entre pares de objetos.

2.2 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 [7]. 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 notación $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ para indicar el conjunto total de parámetros de un HMM.

El enfoque típico para realizar la clasificación cuando se usan HMMs adopta el criterio de máxima verosimilitud, que consiste en que una secuencia desconocida \mathbf{O} se le asigna a la clase que muestre la máxima verosimilitud $P(\mathbf{O}/\lambda)$. Lo anterior requiere el entrenamiento de L HMMs para un problema de L clases. En el enfoque basado en proximidades, en lugar de entrenar un HMM por cada clase, se entrena un modelo por cada secuencia \mathbf{O} de entrenamiento.

2.3 Transformación de las matrices de los HMM

2.3.1 Definición de distancia

La cuestión básica de la estrategia basada en disimilaridad está en cómo definir las en el contexto de un HMM. Dado un HMM λ y una secuencia \mathbf{O} , existe un método estándar (el procedimiento de forward-backward) para calcular $P(\mathbf{O}_i/\lambda_j)$, es decir, la probabilidad de que la secuencia \mathbf{O}_i fue generada por el modelo λ_j . Esta medida cuantifica la verosimilitud y mide qué tan bien la secuencia se acomoda al modelo. Una opción natural consiste en definir la similitud $d_{ij} = d(\mathbf{O}_i, \mathbf{O}_j)$ entre dos secuencias \mathbf{O}_i y \mathbf{O}_j como [5]

$$d_{ij} = d(\mathbf{O}_i, \mathbf{O}_j) = \frac{\log P(\mathbf{O}_i | \lambda_j)}{T_i} \quad (1)$$

donde λ_j es el HMM entrenado con la secuencia \mathbf{O}_i y T_i es la longitud de la secuencia \mathbf{O}_i . El factor $1/T_i$ es el factor de normalización introducido para tener en cuenta secuencias de diferente longitud. La medida de similitud definida por (1) no es simétrica. En este trabajo empleamos la distancia entre un par de HMMs

definida en [8] que cumple las condiciones de simetría. La distancia entre dos HMMs está definida como

$$d(\mathbf{O}_1, \mathbf{O}_2) = -\log \sigma(\mathbf{O}_1, \mathbf{O}_2) \quad (2)$$

donde $\sigma(\mathbf{O}_1, \mathbf{O}_2) = \sqrt{P_{21}P_{12}/P_{11}P_{22}}$ se conoce como medida de similaridad, toma valores entre 0 y 1 mientras $P_{ij} = P(\mathbf{O}_i/\lambda_j)^{1/T_i}$.

2.3.2 Matrices de los HMM en el espacio de disimilaridad

Para transformar los parámetros λ al espacio de disimilaridad, considérese un problema de clasificación de L clases. Para cada clase $\ell \in \{1, 2, \dots, L\}$ se tiene un conjunto N_ℓ de secuencias de entrenamiento $S_\ell = \{\mathbf{O}_1^\ell, \mathbf{O}_2^\ell, \dots, \mathbf{O}_{N_\ell}^\ell\}$. Así, $N = \sum_\ell N_\ell$ es el tamaño total del conjunto de entrenamiento $S = \bigcup_{\ell=1}^L S_\ell$. Sea $\mathfrak{R} = \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_R\}$ un conjunto de R prototipos u objetos representativos; estos objetos pueden pertenecer a un conjunto de secuencias de entrenamiento ($\mathfrak{R} \subseteq S$) o pueden definirse por algún procedimiento como se muestra más adelante. Ahora, sea $D_{\mathfrak{R}}(\mathbf{O})$ una función que retorna el vector de disimilaridades entre una secuencia arbitraria y todas las secuencias en \mathfrak{R} , así (usando la ecuación (2)),

$$D_{\mathfrak{R}}(\mathbf{O}) = \begin{bmatrix} d(\mathbf{O}, \mathbf{P}_1) \\ \vdots \\ d(\mathbf{O}, \mathbf{P}_R) \end{bmatrix} \in \mathbb{R}^R \quad (3)$$

Una vez se ha definido el espacio de disimilaridades \mathbb{R}^R , cualquier clasificador puede usarse en principio. Dependiendo de la elección de \mathfrak{R} se pueden adoptar diferentes enfoques; la opción básica consiste en elegir $\mathfrak{R} = S$, el conjunto total de observaciones de entrenamiento. Con esta opción, la dimensión del espacio de disimilaridad es igual a N , la cardinalidad del conjunto de entrenamiento. En este caso, se debe diseñar un clasificador en un espacio N -dimensional usando solamente N secuencias de entrenamiento, por lo cual el clasificador puede quedar deficientemente entrenado. En este trabajo se adoptan diferentes enfoques de selección de prototipos para escoger \mathfrak{R} , a fin de disminuir la dimensionalidad del espacio de entrenamiento y el número de disimilaridades que deben calcularse para la clasificación de una nueva secuencia.

2.4 Selección de prototipos

Para la selección del conjunto \mathfrak{R} se comparan los siguientes métodos [6]:

I. AleatoriaC. Se realiza una selección aleatoria de r secuencias por clase, de forma tal que $\ell r = R$.
Aleatoria. Se realiza una selección aleatoria de R secuencias a partir de las N secuencias originales.

II. R-centrosC. Esta técnica se aplica separadamente sobre cada clase. Para cada clase ℓ , se busca escoger r secuencias que queden igualmente distribuidas con respecto a la información de la matriz de disimilares construida con las secuencias S_ℓ . Los pasos del algoritmo son

A. Se selecciona un conjunto inicial $\mathfrak{R}_\ell = \{\mathbf{P}_1^\ell, \mathbf{P}_2^\ell, \dots, \mathbf{P}_r^\ell\}$ escogido aleatoriamente a partir de S_ℓ .

B. Para cada \mathbf{O}_i^ℓ que pertenece a S_ℓ se encuentra su vecino más cercano en \mathfrak{R}_ℓ . Sea J_j , $j = 1, 2, \dots, r$ sea un subconjunto de S_ℓ que consiste en las secuencias que tiene el mismo vecino más cercano \mathbf{P}_j^ℓ en \mathfrak{R}_ℓ .

C. Para cada J_j se encuentra su centro c_j , que es la observación para la cual la máxima distancia a todos los otros objetos en J_j es mínima.

D. Para cada centro c_j , si $c_j \neq \mathbf{P}_j^\ell$, se reemplaza \mathbf{P}_j^ℓ por c_j en \mathfrak{R}_ℓ . Si se hace cualquier reemplazo se regresa al punto B. De otra manera, se para el procedimiento. La representación final \mathfrak{R} consiste de todos los subconjuntos \mathfrak{R}_ℓ .

III. R-centros. Se realiza el procedimiento R-centrosC sobre todo el conjunto S .

3. RESULTADOS

3.1 Aplicación a la identificación automatizada de fuentes de espigas

La aplicación de la metodología propuesta se emplea para la identificación automatizada de fuentes de espigas. A continuación se describe brevemente en qué consiste este problema. 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) [9]. 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 microestimulador, 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 [10]. 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 problema de reconocimiento de patrones consiste,

entonces, en determinar con base en registros de la actividad captada por microelectrodos, la localización exacta del microelectrodo en el cerebro. En particular, se requiere clasificar entre dos zonas conocidas como tálamo y subtálamo. La figura 2 resume la metodología descrita



Figura 2. Descripción completa de la metodología

3.2 Configuración del experimento

3.2.1 Base de datos

La base de datos pertenece a la Universidad Politécnica de Valencia. Incluye 190 muestras de las zonas tálamo y 95 de la zona subtálamo de 5 pacientes diferentes. Cada señal fue muestreada a una frecuencia de 24 KHz y las señales tienen longitudes de 10 segs.

3.2.2 Extracción de características

Las características extraídas se obtienen usando la transformada wavelet discreta [11]. Cada señal es segmentada en ventanas de longitud igual a 1 s con un traslape de 500 ms. Para cada marco se aplica la transformada wavelet Haar, debido a que su mejor resolución en el tiempo es apropiada para describir con mayor precisión las características de la señal. Por cada una de 7 escalas de detalle se selecciona el coeficiente con la amplitud mayor para ser utilizado como parte del vector de observación.

3.2.3 Estructura 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 [1] con tres, cinco y diez estados (N_S) y 64, 128 y 256 palabras (N_W) para el libro de códigos [1, 7]. El grupo de parámetros λ que describe el modelo HMM de cada tipo de zona se estima utilizando el algoritmo estándar de Maximización de la Esperanza (EM Expectation-Maximization) [7].

3.2.4 Validación

Se comparan los resultados de usar el enfoque tradicional basado en máxima verosimilitud (ML), el enfoque basado

en análisis de proximidad con el espacio completo y con la selección de prototipos. Se usan dos clasificadores paramétricos: un clasificador lineal basado en una densidad normal regularizada (RLNC) y un clasificador cuadrático basado en una densidad normal regularizada (RQNC) [3] y dos clasificadores no paramétricos: un clasificador de Parzen optimizado (PARZEN) y un clasificador de vecino más cercano (1-NN) [12]. El sistema se entrena variando el número de prototipos desde 6 hasta 70 y se repite 5 veces debido a que la inicialización es aleatoria. La validación se hace con el conjunto de observaciones no utilizadas para el entrenamiento, esto es, conjuntos desde 249 hasta 215 elementos. En los resultados sólo se muestran los errores correspondientes al sistema construido con 70 prototipos ya que luego de pruebas preliminares se estableció que con esta cantidad, el error de los clasificadores arriba mencionados se estabiliza a un punto relativamente fijo.

3.3 Experimento

En las tablas 1, 2, 3, 4 y 5 se muestran los resultados obtenidos para los cuatro métodos de selección de prototipos descritos en la sección 5 y los cuatro clasificadores mencionados en la sección anterior. En cada caso, se presenta el menor error de clasificación obtenido para alguno de los valores de N_S y N_W , parámetros del HMM. Además se incluyen las desviaciones estándar del error producto de la repetición del experimento.

	N_S	N_W	Error
ML	10	64	9.20 ± 6.40

Tabla 1. Resultados obtenidos utilizando el enfoque tradicional con ML

El resultado mostrado en la tabla 1, obtenido con el enfoque tradicional basado en máxima verosimilitud es satisfactorio, sin embargo la desviación estándar del error es muy grande lo cual implica que el clasificador es bastante sensible a las observaciones utilizadas para el entrenamiento. Esto último justifica la necesidad de seleccionar adecuadamente el conjunto de entrenamiento para obtener tasas de error más estables y por consiguiente clasificadores más consistentes.

	N_S	N_W	Error
RLNC	5	64	9.48 ± 1.82
RQNC	5	64	11.01 ± 2.76
1-NN	10	128	31.59 ± 1.20
PARZEN	3	64	13.67 ± 2.90

Tabla 2. Resultados obtenidos utilizando *AleatoriaC* para la selección de prototipos

Los resultados de la tabla 2 muestran que para la selección de prototipos utilizando el método *AleatoriaC* y el clasificador RLNC el error alcanzado es menor,

incluso con una desviación estándar menor que la obtenida por cualquiera de los otros tres clasificadores considerados para el espacio de disimilaridad y a su vez mucho menor que la obtenida por el clasificador usando el enfoque tradicional con ML.

	N_S	N_W	Error
RLNC	3	256	9.71 ± 0.60
RQNC	10	256	11.77 ± 1.45
1_NN	5	64	32.02 ± 2.24
PARZEN	5	128	12.15 ± 1.74

Tabla 3. Resultados obtenidos utilizando *Aleatoria* para la selección de prototipos

En este caso, para el método *Aleatoria*, los resultados de la tabla 3 muestran un mejor desempeño por parte de RLNC de la misma manera que para el método anterior pero con un incremento en el error de al rededor 0.3%. De otro lado, la desviación estándar obtenida con respecto al experimento anterior y RLNC es mucho menor e inferior al 1%.

	N_S	N_W	Error
RLNC	10	64	8.91 ± 0.92
RQNC	10	128	7.91 ± 1.36
1-NN	3	64	27.25 ± 0.33
PARZEN	10	64	13.25 ± 3.59

Tabla 4. Resultados obtenidos utilizando *R-CentrosC* para la selección de prototipos

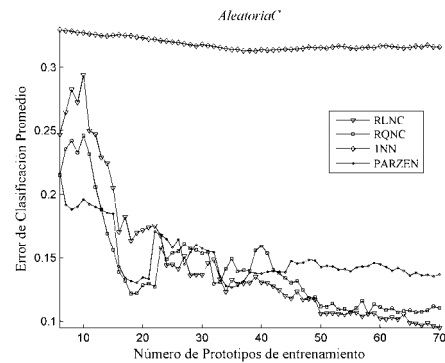
Para el caso de *R-CentrosC*, como se muestra en la tabla 4 los resultados obtenidos muestran que el desempeño del clasificados RQNC es superior, con una desviación estándar considerablemente pequeña aunque mayor a la obtenida por RLNC.

	N_S	N_W	Error
RLNC	3	128	8.71 ± 1.17
RQNC	5	256	10.41 ± 0.77
1-NN	10	128	33.06 ± 1.70
PARZEN	10	256	12.73 ± 1.06

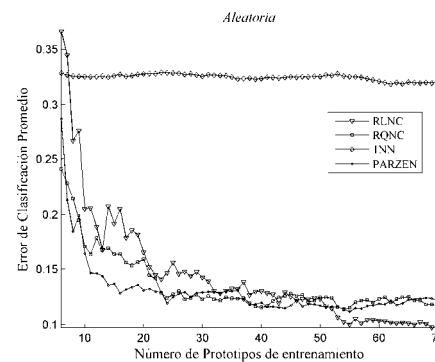
Tabla 5. Resultados obtenidos utilizando *R-Centros* para la selección de prototipos

En la tabla 5 se muestran los resultados obtenidos para el método *R-Centros*, el cual evidencia de nuevo, que el error obtenido por RLNC es menor en comparación a los otros clasificadores tenidos en cuenta en este trabajo. Por último, en las figura 3 se muestran gráficas de error de entrenamiento con respecto al número de prototipos empleados, para los cuatro métodos de selección de prototipos y cada uno de los clasificadores construidos sobre el espacio de disimilaridad.

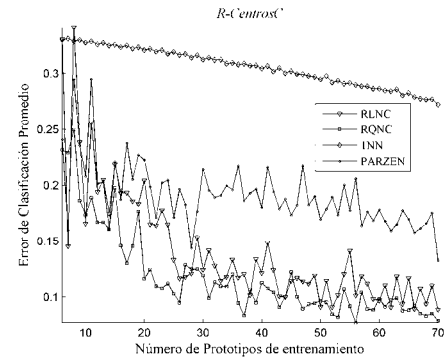
En la figura 3 se puede observar que si bien las tasas de error de clasificación utilizando *R-Centros* y *R-CentrosC* son menores, *Aleatoria* y *AleatoriaC* muestran un mejor



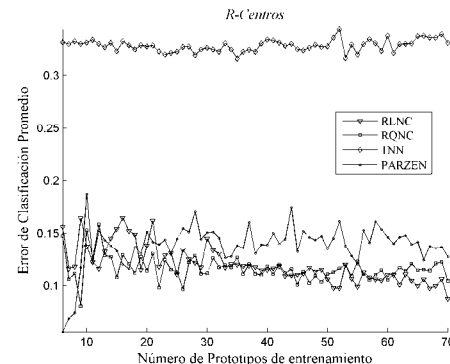
(a) Resultados con *AleatoriaC*



(b) Resultados con *Aleatoria*



(c) Resultados con *R-CentrosC*



(d) Resultados con *R-Centros*

Figura 3. Resultados obtenidos variando la cantidad de prototipos empleados para el entrenamiento

desempeño en términos de suavidad en el decremento del error a causa del aumento de la cantidad de prototipos en el conjunto de entrenamiento.

4. DISCUSIÓN

La identificación de fuentes de espigas en el espacio de disimilaridades de modelos ocultos de Markov mediante la selección de prototipos permite obtener altos porcentajes de acierto. Los resultados obtenidos evidencian que el mejor clasificador es RQNC, en términos de error y desviación estándar, como una medida de la estabilidad del clasificador. Sin embargo, RLNC muestra ser menos sensible a la selección de prototipos considerando que en todos los casos obtuvo tasas de error similares y en tres casos superiores a RQNC, con desviaciones estándar inferiores. Además, que los métodos de selección de prototipos que no hacen la selección puramente aleatoria, esto es, *R-Centros* y *R-CentrosC*, producen mejores resultados de clasificación, inferiores al valor alcanzado usando el enfoque tradicional con ML, pero con desviaciones estándar considerablemente más bajas. Resumiendo, los parámetros finales del sistema para el cual el error es menor son: $N_S = 10$, $N_W = 128$, selección de prototipos con *R-CentrosC* y clasificador RQNC. Sin embargo, utilizando *R-Centros* o *R-CentrosC* en conjunto con RLNC es todavía posible obtener una tasa de error inferior al 9% con desviaciones estándar menores y en uno de los casos inferior al 1%.

Usando el espacio de disimilaridades es posible aumentar el porcentaje de acierto sobre la regla de clasificación MAP. El porcentaje de acierto máximo se mantiene usando únicamente 70 prototipos de la base de datos, esto es, más o menos un 25% de la cantidad de observaciones totales.

Como trabajo futuro se propone cambiar los clasificadores, por ejemplo, mediante clasificadores basados en máquinas de soporte vectorial. Además de emplear criterios para hacer de manera automatizada selección del modelo en el HMM, a saber, el número de estados y el número de palabras de código.

5. AGRADECIMIENTOS

Este trabajo es financiado por Colciencias bajo el contrato 1110-14-17904.

6. BIBLIOGRAFÍA

- [1] L. R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*. Vol. 77. 1989. pp 257-86.
- [2] A. Cohen. Hidden Markov models in biomedical signal processing. En *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 3. 1998. 1145-50
- [3] E. Pekalska, R. Duin. Dissimilarity representations allow for building good classifiers. *Pattern Recognition Letters*. Vol. 23. 2002. 943-56.
- [4] M- Álvarez, J.D. Arias, G. Castellanos, J.J. Godino. Identificación de voces disfuncionales usando HMM y análisis de proximidad. En *Libro de actas del XXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica, CASEIB 2005*. 2005 pp. 125-128
- [5] M. Bicego, V. Murino, M. Figueredo. Similarity-based classification of sequences using hidden Markov models. *Pattern Recognition*. Vol. 37. 2004. pp. 2281-91.
- [6] E. Pekalska, R. Duin, P. Paclík. Prototype selection for dissimilarity based classifiers. *Pattern Recognition*. Vol. 39. 2006. pp. 189-208.
- [7] X. Huang, A. Acero, H. W. Hon. *Spoken Language Processing*. Prentice Hall. Upper Saddle River, New Jersey. 2001. pp
- [8] M. C. Nechyba, Y. Xu. Stochastic Similarity for Validating Human Control Strategy Models. *IEEE Transactions On Robotics And Automation*. Vol. 14. 1998. 437-51.
- [9] J.H. Falkenberg, J. McNames, K.J. Burchiel. Statistical methods of analysis and visualization of extracellular microelectrode recordings. En *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology*. Vol 3. 2003. pp 2515-18
- [10] R.A. Santiago, J. MacNames, K. Burchiel, George G. Lendaris. Developments in understanding neuronal spike trains and functional specializations in brain regions. *Neural Networks*. Vol 16. 2003. pp. 601-07.
- [11] A.A. Orozco, C. Guarnizo, M. Álvarez, G. Castellanos, E. Guijarro. Spike Source Identification using Artificial Intelligence Techniques. En *Proceedings of The 3RD IEE International Seminar on Medical Applications of Signal Processing*. Londres. 2005. pp. 105-109.
- [12] R. Duda, P.E. Hart. D.G. Stork. *Pattern Classification*, Segunda ed. John Wiley & Sons, INC, 2001.