MODELOS DE MARKOV Y CUANTIFICACION VECTORIAL POR MEDIO DE REDES DE KOHONEN.

E.Monte y J.B.Mariño.
Dpt. de Teoría de la Señal y Comunicaciones. ETSITB.UPC.
Apartado 30.002. Barcelona

ABSTRACT

In this paper we present a speech recognition system based on Hidden Markov Models, which uses as a quantifier a fonotropic map. This kind of arquitecture has been used for tasks like discriminating the phonemes of the Japanese language /1/. We will use a similar arquitecture in order to construct a system that might be able to recognize isolated words (digits) independently of the speaker and we will compare the results with those of a classical system (RAMSES) /7/ where the codebook is trained using the LBG algorithm /6/.

INTRODUCCION

Los modelos ocultos de Markov son una técnica muy extendida en el campo del reconocimiento del habla /1/. En este artículo estudiaremos las prestaciones de un sistema de reconocimiento del habla en el que se integra un cuantificador vectorial basado en los mapas fonotrópicos /2/, /3/, /4/ y compararemos las prestaciones del sistema con otro en el que el entrenamiento de los modelos se realiza mediante el algoritmo LBG /6/.

El objetivo del cuantificador vectorial consiste en realizar una compresión de los datos de tal manera que se sustituye un vector por una etiqueta que puede ser un número entero. Sin embargo al realizar la cuantificación vectorial mediante el mapa fonotrópico, nos encontramos que además aparece una propiedad de colindancia que le da una forma especial a la matriz de probabilidades de emisión del modelo oculto de Markov. Esta propiedad está explicada en /8/ y es muy interesante, pues demuestra que realizar la cuantificación vectorial mediante el mapa fonotrópico proporciona información adicional que no se obtiene al realizar la cuantificación mediante otros métodos. En todo caso en este artículo estudiaremos únicamente las prestaciones de un sistema de reconocimiento usando el mapa fonotrópico como si fuera un cuantificador convencional.

MAPAS FONOTROPICOS.

El algoritmo de entrenamiento de los mapas fonotrópicos, también conocido como LVQ (Learning Vector Quantization), está explicado con detalle en las referencias /3/, /4/. Se trata de un

algoritmo adaptativo que aprende de manera iterativa la posición de los centroides, con la propiedad de que ordena estos centroides sobre un plano siguiendo como criterio de orden una proyección no ortogonal del espacio de características sobre un plano (mapa fonotrópico). Esta proyección genera la propiedad que llamamos de colindancia, que consiste en que centroides colindantes en el espacio de las características son colindantes sobre el mapa fonotrópico. Por lo que este tipo de cuantificación vectorial además de proporcionar información sobre cual es el centroide más cercano a un punto dado en el espacio de las características, además nos proporciona información sobre los centroides de su entorno (colindantes).

EXPERIMENTOS REALIZADOS.

Los experimentos se realizaron con una base de datos formada por los dígitos catalanes. La base de datos tiene las características siguientes:

Número de locutores: 10 (seis hombres y cuatro mujeres).
Corpus: 10 palabras (los dígitos del cero al nueve)
Repeticiones de cada dígito: 10 repeticiones por dígito.

Para comparar las prestaciones de la cuantificación mediante los mapas Fonotrópicos con el algoritmo LBG, hicimos varios experimentos.

PRIMER EXPERIMENTO: Comparamos el comportamiento de ambos algoritmos cuando el reconocimiento se realiza dependiendo del locutor (los locutores de reconocimiento son los mismos que los de entrenamiento) y cuando se realiza independiente del locutor (los locutores de reconocimiento son diferentes a los locutores de entrenamiento). Las estadísticas las calculamos en los dos casos de la manera siguiente:

Dependiente del locutor: Reconocemos con los 1000 elementos de la base de datos.
Independiente del locutor: Entrenamos con 9 locutores y reconocemos con el locutor restante, y rotamos 10 veces hasta conseguir reconocer todos los locutores de la base de datos.

Hemos observado que ambos sistemas se comportan de manera similar en cuanto a tasa de error cuando el sistema funciona en modo dependiente del locutor, dando resultados muy similares para diversos tamaños del codebook. Sin embargo al realizar el reconocimiento de manera independiente del locutor se observa una marcada diferencia en cuanto a la tasa de reconocimiento. Al realizar la cuantificación por el método LVQ la tasa de error se reduce en un 50% (pasa de 6,7 % del LBG al 3,3 % del LVQ). Esta mejora es debida a la propiedad de colindancia (a la que T. Kohonen denomina topologica) pues el algoritmo de entrenamiento no sólo calcula los

centroides tomando en cuenta el centro de masas, sino que además toma en cuenta la distribución relativa de los vecinos en el espacio de características.

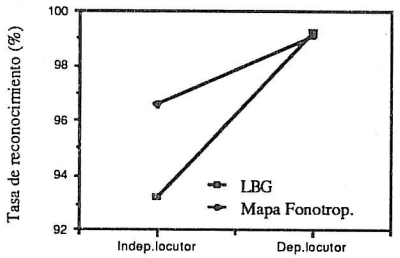

Figura 2. Resultados de reconocimiento usando como cuantificadores vectoriales los algoritmos LVQ (Mapa Fonotrópico) y el algoritmo LBG. (Codebook de dimensión 64)

SEGUNDO EXPERIMENTO: consisitió en comparar el comportamiento del algoritmo LBG con el Mapa de Kohonen para el caso de reconocimiento independiente del locutor para varios tamaños de codebooks. En la figura 3 se muestran los resultados de reconocimiento y se observa que para el caso independiente del locutor el sistema que usa el mapa fonotrópico se comporta de una manera marcadamente mejor que el que usa el LBG para una gama amplia de tamaños del codebook. La explicación es la misma que en el caso anterior y está relacionada con el hecho de que el mapa fonotrópico proporciona información adicional que no se obtiene al usar el algoritmo LBG.

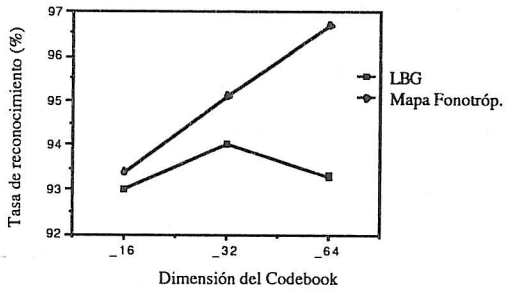

Figura 3. Tasa de reconocimiento al variar la dimensión del codebook de 16 hasta 64, para los dos algoritmo comparados.

COMENTARIOS SOBRE LOS MAPAS FONOTROPICOS.

Una de las diferencias importantes entre el algoritmo LBG y el LVQ de cara a una implementación correcta es la cantidad de grados

de libertad que tiene el LVQ a la hora de realizar el entrenamiento de codebook. En la figura 4 se muestran los resultados de reconocimiento que se obtiene al variar el parámetro de adaptación. Aunque no se muestra en la gráfica hemos encontrada que el punto de inflexión de la curva de reconocimiento (valor del parámetro de adaptación para el que se obtiene un error de reconocimiento mínimo) se encuentra en 1.0, al usar 2.0 la tasa de error empeora de tal manera que decidimos que no valía la pena realizar los mil reconocimientos para tener una estadística fiable. La razón por la que incluimos esta gráfica es para hacer patente las grandes diferencias en cuanto a resultados que se pueden obtener al modificar alguno de los grados de libertad del algoritmo.

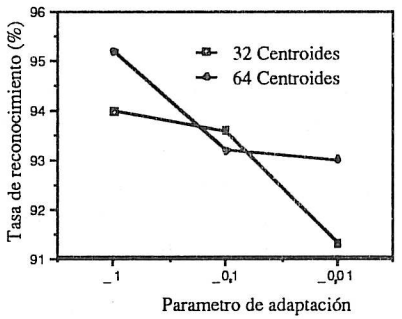

Figura 4. Variación de la tasa de reconocimiento en función del paso de adaptación del algoritmo LVQ.

Un aspecto digno de mención relacionado con el uso de la cuantificación basada en los mapas fonotrópicos es la relativa independencia de la tasa de error respecto a la distorsión media del codebook, tal como se puede ver comparando las figuras 4 y 5. Este aspecto está comentado en la literatura /1/, aunque no explicado de manera rigurosa. En todo caso posiblemente esté relacionado con la propiedad de colindancia ya comentada anteriormente. Los centroides que proporciona este algoritmo tienen la propiedad de estar ordenados por colindancia y aunque no usemos esta información de manera explicita, se encuentra de manera implícita en la distribución de centroides. De todos modos es importante hacer notar que esta independencia de la tasa de reconocimiento respecto a la distorsión media del codebook es únicamente valida cuando los codebooks están bien entrenados, pues se comprueba de manera inmediata que la tasa de reconocimiento es inferior y la distorsión mayor, cuando el codebook se entrena de manera insuficiente.

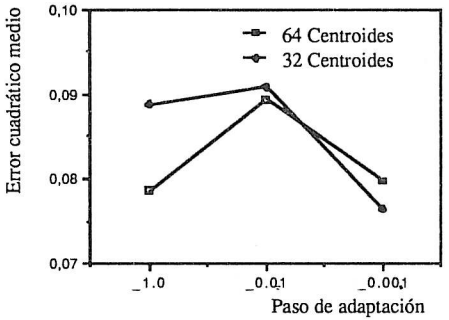

Figura 5. Evolución de la distorsión del cuantificador en función del paso de adaptación.

CONCLUSIONES

En este articulo presentamos una comparación entre dos sistemas de reconocimiento del habla basados en modelos ocultos de Markov. Se observa a partir de los resultados de reconocimiento que el sistema que usa el cuantificador basado en el mapa fonotrópico es superior al sistema basado en el algoritmo LBG. Esta diferencia es patente en el caso de reconocimiento del habla independiente del locutor y esta mejora es debida al hecho de que el los mapas fonotrópicos además de proporcionar información sobre el centroide más cercano, ordena los centroides por colindancia en el espacio de las características.

REFERENCIAS.

/1/ H.Iwamida, S.Katagiri, E.McDermott, y Y.Tohkura. "A Hibrid Speech Recognition System using HMMS with an LVQ-trained Codebook". ICASSP-90.

/2/Lawrence R.Rabiner. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". Proceedings of the IEEE, vol.77, Feb.89.

/3/T.Kohonen. "Self-Organization and Associative Memory". Springer-Verlag.89.

/4/T.Kohonen, "The Neural Phonetic Typewriter". IEEE Computer, Vol. 21,No3, Marzo 88.

/5/ T.Kohonen, K. Torkola, M. Shazakai, J. Kangas, O.Venta."Phonetic Typewriter for Finnish and Japanese". pp 607 ICASSP 88.

/6/Linde, Buzo, Gray."An Algorithm for Vector Quantizer Design". IEEE Trans. Commun., COM-28,1,pp84-95.1980.

/7/ J.B.Mariño, C.Nadeu,A.Moreno,E.Lleida, E.Monte,A.Bonafonte.
"RAMSES: a Spanish demisyllable based continuous speech recognition system". NATO ASI. speech Recognition and Understanding: Recent Advances, Trends and Applications, Cetraro Italy -90

/8/ E.Monte, J.B.Mariño, E.LLeida. "Alisado de los Modelos De Markov mediante Redes Fonotrópicas. Alicación al Reconocimiento del Habla."
SEPLN-91.