

Reconeixement dels Dígits Catalans Utilitzant Models de Markov Continus

Sara Garrigosa Rivas, A. Moreno.

Departament de Teoria del Senyal i Comunicació.

En aquest projecte s'apliquen les tècniques de reconeixement de la parla aïllada basades en els Models de Markov Ocults (HMM), però reemplaçant la representació dels símbols discrets dels senyals de veu per funcions densitat de probabilitat Gaussians contínues. D'aquesta manera, eliminem l'error de quantificació introduït per la representació discreta. El vocabulari sobre el qual es realitzaran les diferents proves està compost pels 10 dígits catalans (del 0 al 9). Observarem que els resultats obtinguts són millors que en el cas de HMM discrets.

I. INTRODUCCIO.

Un HMM (Hidden Markov Model), és un procés estocàstic doble on un d'ells queda ocult (Hidden) i només pot ser conegut a través de l'altre procés estocàstic produït per la seqüència de símbols observats [1].

Suposem que el senyal a representar pel HMM consisteix en una seqüència de vectors d'observacions $O = \{O_1, O_2, \dots, O_T\}$ on cada O_i és un vector que caracteritza el senyal en el temps $t=i$, i on O_i , en el cas discret, ha sigut substituït, via algun tipus de VQ, per algun dels S símbols. D'aquesta manera un HMM discret tal com el de la figura 1 queda caracteritzat pels següents paràmetres:

N: nombre d'estats del model

A=[a_{ij}], 1 ≤ i, j ≤ N: matriu de transició d'estats on a_{ij} és la probabilitat de fer una transició de l'estat i a l'estat j. En el nostre cas imposem la restricció: a_{ij}=0 j < i, j > i+2.

B=[B_j(k)], 1 ≤ j ≤ N, 1 ≤ k ≤ S: funció de probabilitat d'observació,

on b_{jk} és la probabilitat d'observar el símbol k a l'estat j.

Així tenim que la probabilitat de que la seqüència d'observacions O hagi estat generada pel model l és:

$$P(O/\lambda) = \sum_{i_1, \dots, i_T} \pi_{i_1} B_{i_1}(O_1) a_{i_1, i_2} B_{i_2}(O_2) \dots a_{i_{T-1}, i_T} B_{i_T}(O_T)$$

on π_{i1} és la probabilitat de que inicialment estem a l'estat i1

En la figura 1 es mostra el tipus de HMM considerat aquí. Es tracta d'una cadena de Markov d'esquerra a dreta que comença a l'estat 1 i acaba a l'estat N. La seqüència d'estats no és observable.

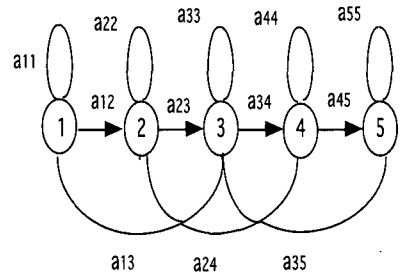


FIG. 1: Representació d'un HMM per N=5 estats

II. HMM CONTINUS.

En el cas continu els O_i ja no són substituïts pels símbols. Ara tenim les funcions densitat de probabilitat $B = \{B_j(x)\}$, 1 ≤ j ≤ N on $B_j(x)dx$ és la probabilitat de que el vector O_i estigui entre x i dx. Nosaltres considerarem funcions Gaussians:

$$B_j(x) = \sum_{k=1}^M c_{jk} f(x, \mu_{jk}, U_{jk})$$

on $f(x, \mu, U)$ és un funció densitat de dimensió D (D=12 en el nostre cas) de mitja μ i de covariança U. Llavors els paràmetres que caracteritzen el nostre HMM continu són:

N: nombre d'estats.

M: nombre de mescles (o f.d.p. per estat).

D: dimensió dels vectors.

A=[a_{ij}]: matriu de transició d'estats.

C=[c_{jk}]: matriu de guany de la mescla.

μ=[μ_{jk}]: vector de les mitges de les components de les mescles.

U=[U_{jkde}]: matrius de les covariances de les components de les mescles.

2.1. HMM AMB INFORMACIO DE L'ESPECTRE, VARIACIO ESPECTRAL I ENERGIA DIFERENCIAL.

Per poder treballar amb el senyal l'hem hagut de reduir prèviament a símbols a través de l'anàlisi LPC (Linear Predictive Coding). Aquest ens proporciona un vector, de dimensió 12 en el nostre cas, que conté la informació espectral de la trama que estem estudiant dintre del nostre senyal de veu.

L'objectiu del nostre projecte és implementar un sistema en el qual els models de Markov utilitzin, a més a més dels coeficients cepstrum, d'altres informacions per tal de poder millorar els resultats.

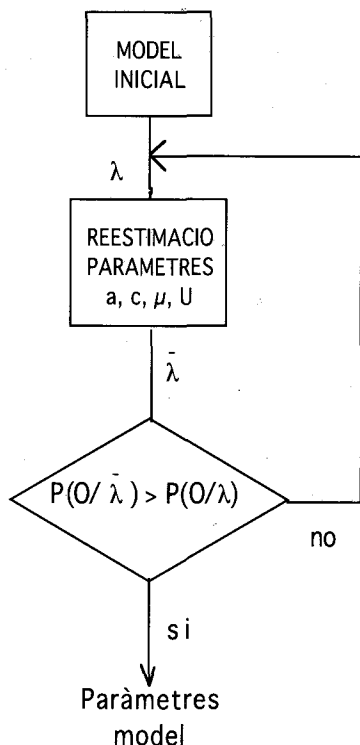
Furui [2] va proposar la utilització de la variació temporal tant

de l'espectre com de l'energia. Això va donar lloc als coeficients cepstrals diferenciats, ΔC , i a l'energia diferencial, ΔE . Introduint aquestes tres informacions ens queda que

$$B_j(x_t) = B_j(x_t^c, x_t^{dc}, x_t^{de}) = B_j(x_t^c) B_j(x_t^{dc}) B_j(x_t^{de}) = \prod_{nc=1}^{ncob} \sum_{k=1}^{M(nc)} f_k(x_t^{nc})$$

III. FASE D'ENTRENAMENT DEL MODEL

Si suposem que $Pr(O/l)$ és la probabilitat de que la seqüència d'entrenament hagi estat generada pel model l , aquesta fase consistirà en ajustar els paràmetres del nostre model per tal de maximitzar aquesta probabilitat. Els passos a seguir són:



L'algorisme utilitzat en la reestimació dels paràmetres és l'implementat per Baum-Welch [1] i [3]. Però de fet les proves realitzades amb aquest procediment de reestimació ens mostren que les mitges, μ , són els paràmetres més sensibles a la inicialització del model.

Així si comparem els resultats obtinguts inicialitzant el model a partir dels models discrets [4] amb els resultats obtinguts amb una inicialització per Viterbi [1] i [5], podem observar que aquests últims són millors.

IV. RESULTATS

Per les proves tenim 100 versions de cada dígit dites per 10 locutors. Els resultats que es mostren corresponen a proves realitzades entrenant només 8 dels locutors i reconeixent els 2 que han estat fora de l'entrenament.

En aquesta primera taula tenim els resultats obtinguts entrenant els models amb HMM continus on la primera columna correspon als dígit, la segona correspon al tant per cent d'error quan s'inicialitza el model amb el codebooks discrets i la tercera quan la inicialització ha sigut per Viterbi. En el primer cas l'error total és del 7% i en el segon cas del 1'7%.

	Inicialització discrets	Inicialització Viterbi
0	6%	0%
1	30%	3%
2	1%	0%
3	8%	1%
4	0%	0%
5	4%	1%
6	20%	4%
7	1%	0%
8	0%	1%
9	0%	1%

taula 1

En aquesta segona taula comparem els resultats que es tenien en el cas d'utilitzar els HMM discrets amb els nostres. Amb els models discrets l'error total produït és del 3'3% i amb els continus és del 1'7%.

	HMM discrets	HMM continus
0	4%	6%
1	1%	3%
2	1%	0%
3	3%	1%
4	0%	0%
5	3%	1%
6	4%	4%
7	11%	0%
8	5%	1%
9	1%	1%

taula 2

BIBLIOGRAFIA:

- [1] L.R.Rabiner, B.H.Juang, «An Introduction to Hidden Markov Models». IEEE (Gener 1986), pag. 4-16.
- [2] Sadaoki Furui, «Speaker Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum». IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34 (febrer 1986).
- [3] Biing-Hwuang Juang, L.R.Rabiner, «Mixture Autoregressive Hidden Markov Models for Speech Signals», IEEE Transactions on Acoustics, Speech and Signal Processing (desembre 1985), pag. 1404-12.
- [4] Lluís Serra, «Reconeixement de paraules aïllades mitjançant Models de Markov Continus». PFC (juny 1990).
- [5] L.R.Rabiner, B.H.Juang, S.E.Levinson, M.M.Sonndhi, «Recognition of Isolated Digits Using Hidden Markov Models with Continuous Mixture Densities», (novembre 1984), pag 1211-32.