Modele HMM pentru prelucrarea semnalului sonor

Lect.dr. Catalina-Lucia COCIANU
Catedra de Informatica Economica, A.S.E. Bucuresti

The research reported in the paper focused on the HMM modeling in pattern recognition, with application in speech processing. A speech recognizer is a device that automatically transcribes speech into text and can be thought of as a voice-actuated "typewriter". The recognizer is usually based on some finite vocabulary that restricts words that can be printed out. The mathematical formulation of the speech recognizer design problem is based on a statistical approach, involving the hidden Markov model (HMM).
Keywords: *HMM, acoustic modeling, language modeling, speech recognition, vector quantization.*

Model matematic pentru recunoasterea semnalului sonor

Modelul matematic pentru recunoasterea automata a semnalului sonor poate fi descris astfel. Fie A data acustica analizata, definita prin intermediul secventei de simboluri $a_1, a_2, ..., a_m$ din alfabetul $\boldsymbol{A}$, în care indicele i, $1 \le i \le m$, reprezinta momentul de timp la care este generat simbolul a_i $A = a_1, a_2, ..., a_m$, $\forall 1 \le i \le m, a_i \in \boldsymbol{A}$, $W = w_1, w_2, ..., w_n$ o secventa de n cuvinte din vocabularul $\boldsymbol{V}$ si $P(W \mid A)$ probabilitatea ca la pronuntarea datei acustice A sa fie scris sirului de cuvinte W. Dispozitivul de recunoastere automata asociaza datei acustice A observata acea secventa de cuvinte $\hat{W}$ cu proprietatea ca

$$\hat{W} = arg \max_{W} P(W \mid A) \quad (1)$$

Aplicând formula Bayes pentru calculul $P(W \mid A)$ obtinem,

$$P(W \mid A) = \frac{P(W)P(A \mid W)}{P(A)}, \quad (2)$$

unde $P(W)$ este probabilitatea pronuntarii sirului de cuvinte W, $P(A \mid W)$ este probabilitatea ca la pronuntarea sirului de cuvinte W semnalul acustic observat sa fie A si $P(A) = \sum_{W'} P(W')P(A \mid W')$ este probabilitatea medie ca A sa fie observata. Deoarece în relatia (1) A este fixat, din (2) rezulta,

$$\hat{W} = arg \max_{W} P(W)P(A \mid W) \quad (3)$$

Proiectarea dispozitivului de recunoastere a vorbirii este realizata pe baza relatiei (3) si presupune parcurgerea urmatoarelor etape [Jeli, 1997]:

I. *Procesarea acustica*, prezentata schematic în figura 1;

II. *Determinarea unui model acustic pentru calculul valorilor* $P(A \mid W)$; o varianta des utilizata este modelarea HMM (Hidden Markov Model) pentru obtinerea $P(A \mid W)$

III. *Modelarea limbajului*. Aplicarea relatiei (3) pentru calculul $\hat{W}$ presupune calculul probabilitatii a priori $P(W)$; alegerea cea mai utilizata este descompunerea pe baza formulei Bayes,

$$P(W) = \prod_{i=1}^{n} P(w_i \mid w_1, ..., w_{i-1}) \quad (4)$$

În procesul de recunoastere automata a vorbirii, datorita complexitatii ridicate (datorate dimensiunii vocabularului $\boldsymbol{V}$), probabilitatile $P(w_i \mid w_1, ..., w_{i-1})$ pot fi doar estimate. În plus, presupunerea ca alegerea de catre vorbitor a celui de-al i-lea cuvânt depinde de întreaga istorie $w_1, ..., w_{i-1}$ poate fi redusa în general la alegerea lui w_i depinzând de o istorie aleasa într-o clasa de echivalenta $\Phi(w_1, ..., w_{i-1})$ [Ziem, 1995].

Rezulta ca $P(W)$ poate fi calculat prin,

$$P(W) = \prod_{i=1}^{n} P(w_i \mid \Phi(w_1, ..., w_{i-1})) \quad (5)$$

si stabilirea unui model pentru limbajul utilizat revine la determinarea unui clasificator Φ si la selectarea unor metode de estimare a probabilitatilor $P(w_i \mid \Phi(w_1,...,w_{i-1}))$.

IV. Calculul $\hat{W}$ pe baza relatiei (3), prin formularea unor *ipoteze care sa determine restrângerea spatiului de cautare a candidatilor W*.

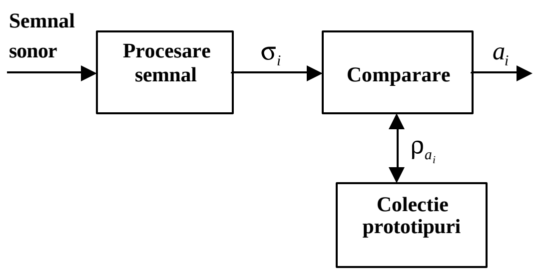

Fig.1. Procesarea acustica

Procesarea de semnal revine la generarea câte unui vector de numere reale σ_i la fiecare interval de timp prestabilit (în general de *s/100*). Componentele vectorului σ_i sunt iesirile unui filtru band-pass aplicat semnalului sonor analizat. Colectia de prototipuri $\Re = \{\rho_1, \rho_2, ..., \rho_K\}$ contine vectori prototip de acelasi tip cu σ_i.

Etapa de comparare presupune calculul celui mai apropiat prototip de σ_i (în sensul distantei euclidiene) si selectarea indexului acelui prototip drept simbolul acustic asociat semnalului sonor transmis. O modalitate de calcul al celui mai apropiat prototip este prin *cuantizare*. Fie L dimensiunea vectorilor prototip si d distanta euclidiana, $d(\sigma,\rho) = \sqrt{\sum_{l=1}^{L} (\sigma(l) - \rho(l))^2}$. Algoritmul de cuantizare (numit si K-means) este descris astfel [Jeli, 1997].

Algoritmul K-means

Pas1. selecteaza numarul de clustere K si numarul de iteratii T;

Pas2. calculeaza $\sigma_i, 1 \le i \le N$, vectorii obtinuti în urma emiterii semnalului sonor pe un interval de timp cunoscut (în practica este utilizat N pentru transmisii de 5 min.);

Pas 3. pentru fiecare vector $\sigma_i, 1 \le i \le N$, este selectat aleator uniform un prototip ρ_j^0, $P(\rho_j^0 = \sigma_i) = \frac{1}{N}$;

Pas 4. partitioneaza $\{\sigma_1, \sigma_2, ..., \sigma_N\}$ în K regiuni $R_j^0, 1 \le j \le K$ definite prin, $\sigma_i \in R_j^0$ daca $d(\sigma_i, \rho_j^0) \le d(\sigma_i, \rho_h^0), \forall h \ne j$;

Pas 5. pentru $t = 1,...,T$:

5.1. calculeaza ρ_j^t centrul fiecarei regiuni $R_j^{t-1}, 1 \le j \le K$, $\rho_j^t = arg\ \min_{\rho} \sum_{\sigma_i \in R_j^{t-1}} d(\sigma_i, \rho)$;

5.2. calculeaza $R_j^t, 1 \le j \le K$ utilizând centrele de greutate ρ_j^t, $\sigma_i \in R_j^t$ daca $d(\sigma_i, \rho_j^t) \le d(\sigma_i, \rho_h^t), \forall h \ne j$.

Modele HMM

Fie $X_1, X_2, \ldots, X_n, \ldots$ o secventa de variabile aleatoare cu valori în $\aleph = \{1,2...,c\}$. În general, pentru orice $n \ge 1$, are loc formula Bayes,

$\forall\ i_1, i_2, ..., i_n \in \aleph$,

$$P(X_1 = i_1, X_2 = i_2, ..., X_n = i_n) = \prod_{j=1}^{n} P(X_j = i_j \mid X_1 = i_1, X_2 = i_2, ..., X_{j-1} = i_{j-1}). \quad (6)$$

Variabilele aleatoare $X_1, X_2, \ldots, X_n, \ldots$ formeaza un *lant Markov* daca, pentru orice $j \ge 1$ si $\forall\ i_1, i_2, ..., i_j \in \aleph$,

$$P(X_j = i_j \mid X_k = i_k, 1 \le k \le j-1) = P(X_j = i_j \mid X_{j-1} = i_{j-1}).$$

Utilizând (6), rezulta,

$\forall\ n \ge 1$ si $\forall\ i_1, i_2, ..., i_n \in \aleph$,

$$P(X_1 = i_1, X_2 = i_2, ..., X_n = i_n) = \prod_{j=1}^{n} P(X_j = i_j \mid X_{j-1} = i_{j-1}). \quad (7)$$

Relatia (7) defineste o clasa de procese stohastice care utilizeaza cantitatea minima de

memorie: valoarea procesului la momentul de timp i depinde numai de valoarea de la momentul i-1 si nu de întreg istoricul procesului.
Un lant Markov se numeste *omogen* daca, pentru orice i>1 si orice $x, x' \in \aleph$,
$P(X_i = x' \mid X_{i-1} = x) = p(x' \mid x)$.
Functia p este numita *functie de tranzitie* si poate fi reprezentata prin intermediul unei matrice $c \times c$ - dimensionala care îndeplineste conditiile:
$\forall x \in \aleph, \sum_{x' \in \aleph} p(x' \mid x) = 1$ si
$\forall x, x' \in \aleph, p(x' \mid x) \geq 0$.
Elementele multimii $\aleph$ se numesc *stari*. Lantul Markov $X_1, X_2, \ldots, X_n, \ldots$ este *un proces cu un numar finit de stari*, tranzitiile dintre stari fiind specificate prin intermediul functiei p. Modelul de lant Markov este *observabil*, în sensul ca, la fiecare moment de timp, iesirea procesului este o multime de stari, fiecare stare fiind corespunzatoare unui eveniment observabil.

Conceptul de HMM

Modelele Markov au proprietatea ca fiecare stare corespunde unui eveniment determinist observabil. Rezulta ca, pentru fiecare stare din model, iesirea este fixata (nu este aleatoare). În scopul eliminarii acestei restrictii, modelul Markov este extins la cazul în care observarea unui eveniment este o functie probabilistica definita pe multimea starilor.
Conceptul de *model Markov ascuns* (**H**idden **M**arkov **M**odel) este definit astfel încât starile lantului Markov genereaza date observabile, în timp ce secventa de stari ramâne ascunsa. Astfel, vor fi definite:

1. alfabetul de iesire $Y = \{0,1,\ldots,b-1\}$;
2. spatiul starilor $\aleph = \{1,2\ldots,c\}$ si s_0 stare initiala unica;
3. distributia de probabilitate a tranzitiilor dintre stari, $p(s' \mid s)$; procesul determinat de multimea starilor si probabilitatea tranzitiilor între stari este un lant Markov;
4. probabilitatea de distributie a iesirilor, asociata fiecarei tranzitii de la o starea s la starea s', $q(y \mid s, s')$, pentru orice $s, s' \in \aleph$; procesul astfel obtinut este un lant Markov.

În cadrul unui HMM, principalele probleme care trebuie rezolvate sunt urmatoarele.

1. Fiind data o secventa observata $\mathbf{y}$ într-un model M, determinati o modalitate eficienta pentru calculul $P(\mathbf{y})$.
2. Fie $\mathbf{y}$ o secventa de iesire observata într-un model M. Calculati cea mai probabila secventa de stari **S** cu proprietatea ca determina generarea sirului **y**.
3. Fie $\mathbf{y}$ o secventa de iesire observata. Determinati modelul Markov ascuns M cu proprietatea ca, în cadrul lui, $P(\mathbf{y})$ este maxima.

În continuare este solutionata prima dintre problemele enuntate mai sus.
Probabilitatea observarii sirului de iesire HMM $y_1, y_2, \ldots, y_k$ este calculata prin:

$$P(y_1, y_2, \ldots, y_k) = \sum_{s_1, s_2, \ldots, s_k} \prod_{i=1}^{k} p(s_i \mid s_{i-1}) q(y_i \mid s_{i-1}, s_i). \qquad (8)$$

O modalitate alternativa de definire a unui HMM presupune evidentierea multimii tranzitiilor de la o stare la alta, fiecarei tranzitii t asociindu-se un simbol de iesire generat cu probabilitate 1 la efectuarea lui t. Fie $L(t)$ si $R(t)$ starea sursa, respectiv starea destinatie a tranzitiei t si $Y(t)$ simbolul de iesire asociat unic lui t. Vom nota cu $p(t)$ probabilitatea ca $L(t)$ sa constituie sursa tranzitiei t: pentru orice $s \in \aleph$, $\sum_{t \mid L(t)=s} p(t) = 1$. Obtinem,

$$p(t) = q(Y(t) \mid L(t), R(t)) p(R(t) \mid L(t))$$

În aceste conditii, relatia (8) este echivalenta cu,

$$P(y_1, y_2, \ldots, y_k) = \sum_{S(y_1, y_2, \ldots, y_k)} \prod_{i=1}^{k} p(t_i), \qquad (9)$$

unde

$$S(y_1, y_2, ..., y_k) = \{t_1, t_2, ..., t_k \mid L(t_1) = s_0, Y(t_i) = y_i, R(t_i) = L(t_{i+1}) \, \forall i = 1, ..., k\}.$$

Calculul probabilitatii $P(y_1, y_2, ..., y_k)$ poate fi realizat recursiv, pe baza notiunii de *latice a drumurilor*. O latice a drumurilor unui HMM evidentiaza evolutia în timp a procesului care a generat secventa $y_1, y_2, ..., y_k$ si este definit ca o multime de siruri de stari elementare, fiecare sir de stari fiind determinat de simbolul de iesire particulara y_i.

Pentru orice stare $s \in \aleph$, sunt definite probabilitatile:

$$\alpha_i(s) = P(y_1, y_2, ..., y_i, s_i = s), \text{ cu } \alpha_0(s) = \begin{cases} 1, & s = s_0 \\ 0, & s \neq s_0 \end{cases}. \quad (10)$$

Rezulta ca $P(y_1, y_2, ..., y_k) = \sum_{s \in \aleph} \alpha_k(s)$, unde probabilitatile α_i sunt calculate recursiv prin,

$$\alpha_i(s) = \sum_{s' \in \aleph} p(y_1, s \mid s') \alpha_{i-1}(s'), \quad (11)$$

$$p(y, s \mid s') \underset{=}{\underline{not.}} q(y \mid s', s) p(s \mid s').$$

Algoritmul Viterbi

Fie $y_1, y_2, ..., y_k$ o secventa de iesire observata. Problema este de a calcula cea mai probabila secventa de stari $s_1, s_2, ..., s_k$ cu proprietatea ca a determinat generarea sirului $y_1, y_2, ..., y_k$. Deoarece

$$P(s_1, s_2, ..., s_k \mid y_1, y_2, ..., y_k, s_0) = \frac{P(s_1, s_2, ..., s_k, y_1, y_2, ..., y_k \mid s_0)}{P(y_1, y_2, ..., y_k \mid s_0)},$$

secventa $s_1, s_2, ..., s_k$ rezulta prin maximizarea $P(s_1, s_2, ..., s_k, y_1, y_2, ..., y_k \mid s_0)$. În plus, datorita faptului ca procesul starilor este Markov, obtinem ca, pentru orice i,

$$P(s_1, s_2, ..., s_i, s_{i+1}, ... s_k, y_1, y_2, ..., y_i, y_{i+1}, ..., y_k \mid s_0) = P(s_1, s_2, ..., s_i, y_1, y_2, ..., y_i \mid s_0) \times$$
$$\times P(s_{i+1}, ... s_k, y_{i+1}, ..., y_k \mid s_i)$$

deci

$$\max_{s_1, s_2, ..., s_k} P(s_1, s_2, ..., s_i, s_{i+1}, ... s_k, y_1, y_2, ..., y_i, y_{i+1}, ..., y_k \mid s_0) =$$
$$= \max_{s_1, s_2, ..., s_{i-1}} P(s_1, s_2, ..., s_{i-1}, s_i, y_1, y_2, ..., y_i \mid s_0) \times \max_{s_i, s_{i+1}, ..., s_k} P(s_{i+1}, ... s_k, y_{i+1}, ..., y_k \mid s_i) \quad (12)$$

Fie $\gamma_i(s_i) = \max_{s_1, s_2, ..., s_{i-1}} P(s_1, s_2, ..., s_{i-1}, s_i, y_1, y_2, ..., y_i \mid s_0)$. Din (12) rezulta,

$$\max_{s_1, s_2, ..., s_k} P(s_1, s_2, ..., s_i, s_{i+1}, ... s_k, y_1, y_2, ..., y_i, y_{i+1}, ..., y_k \mid s_0) =$$
$$= \max_s \left\{ \max_{s_{i+1}, ..., s_k} P(s_{i+1}, ... s_k, y_{i+1}, ..., y_k \mid s) \gamma_i(s) \right\} \quad (13)$$

În consecinta, secventa de stari $s_1, ..., s_{i-1}, s_i, s_{i+1}, ..., s_k$ poate fi calculata prin aplicarea urmatorului algoritm. [Rabi, 1989]

Pentru fiecare s din al i-lea nivel al laticei drumurilor

Pas1. determina $s_1(s), ..., s_{i-1}(s)$, secventa cea mai probabila de stari care a condus la *s*;

Pas2. calculeaza $s_{i+1}(s), ..., s_k(s)$, secventa cea mai probabila de stari care porneste din *s*;

Pas3. calculeaza starea *s* de la nivelul *i* astfel încât secventa completa $s_1(s), ..., s_{i-1}(s), s_i = s, s_{i+1}(s), ..., s_k(s)$ sa fie cu probabilitate maxima.

Algoritmul poate fi definit recursiv, pe baza urmatoarelor observatii.

1. $\gamma_i(s_i) = \max_{s_1, s_2, ..., s_{i-1}} P(s_1, s_2, ..., s_{i-1}, s_i, y_1, y_2, ..., y_i \mid s_0) =$

$$= \max_{s_{i-1}} p(y_i, s_i \mid s_{i-1}) \max_{s_1, s_2, \ldots, s_{i-2}} P(s_1, s_2, \ldots, s_{i-2}, s_{i-1}, y_1, y_2, \ldots, y_{i-1} \mid s_0) =$$
$$= \max_{s_{i-1}} p(y_i, s_i \mid s_{i-1}) \gamma_{i-1}(s_{i-1}) \quad (14)$$
$$2.\ \max_{s_1, s_2, \ldots, s_k} P(s_1, s_2, \ldots, s_k, y_1, y_2, \ldots, y_k \mid s_0) = \max_s \gamma_k(s)$$

Din (14) rezulta algoritmul Viterbi pentru determinarea secventei de probabilitate maxima $s_1, s_2, \ldots, s_k$ prin calculul secventelor de probabilitate maxima $s_1(s), s_2(s), \ldots, s_{i-1}(s), s = s_i$, pentru fiecare stare *s* de la nivelul *i* si selectarea finala (de la nivelul *k*) a unei secvente din multimea $\{s_1(s), s_2(s), \ldots, s_{k-1}(s), s = s_k \,/\, s \in \aleph\}$ care maximizeaza $\gamma_k(s)$. [Jeli, 1997]

Algoritmul Viterbi contine urmatorii pasi:

Pas1. $\gamma_0(s_0) = 1$;

Pas2. pentru toate starile *s* ale primei coloane din laticea drumurilor, calculeaza $\gamma_1(s)$ prin (14), $\gamma_1(s) = \max_{s'} p(y_1, s \mid s')\gamma_0(s') = p(y_1, s \mid s_0)$, $\forall s' \neq s_0, \gamma_0(s') = 0$

Pas3. pentru orice $2 \leq i < k$, calculeaza $\gamma_i(s)$ pe baza relatiei (14), pentru toate starile *s* din cea de-a *i*-a coloana a laticei. Elimina toate tranzitiile din starile *s'* din cea de-a *i*-1-a coloana în starile *s* din cea de-a *i*-a coloana pentru care $\gamma_i(s) > p(y_i, s \mid s')\gamma_{i-1}(s')$. Daca dupa eliminare ramân mai multe tranzitie, selecteaza arbitrar o tranzitie catre *s*.

Pas4. calculeaza starea *s* din cea de-a *k*-a coloana a laticei, care maximizeaza $\gamma_k(s)$. Prin deplasarea inversa de la starea *s* la starea s_0 pe baza tranzitiilor calculate, rezulta secventa de stari de probabilitate maxima, $s_1, \ldots, s_{k-1}, s = s_k$.

Bibliografie

[Char, 1994] Charniak, E., Statistical Language Learning, The MIT Press, 1994

[Gray, 1991] Gray, R., M., Gersho, A., Vector Quantization and Signal Compression, Klewer, Boston, MA, 1991

[Jeli, 1997] Jelinek, F., Statistical Methods for Speech Recognition, The MIT Press, 1997

[Juan, 1992] Juang, B., H, Rabiner, L., R., Issues in Using Hidden Markov Models for Speech Recognition, Advances in Signal Processing, 1992

[Pico, 1994] Picone, J., W., Signal Modeling Techniques in Speech Recognition, Proceedins of IEEE, vol. 81, no. 9, 1994

[Rabi, 1989] Rabiner, L., R., A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, vol. 37, no. 2, 1989

[Youn, 1990] Young, S., J., Competitive Training in Hidden Markov Models, Proceedings of IEEE Intrenational Conference on Acoustic, Speech and Signal Processing, Albuquerque, 1990

[Ziem, 1995] Ziemer, R., E., Tranter, W., H., Principles of Communications, John Wiley, New York, 1995