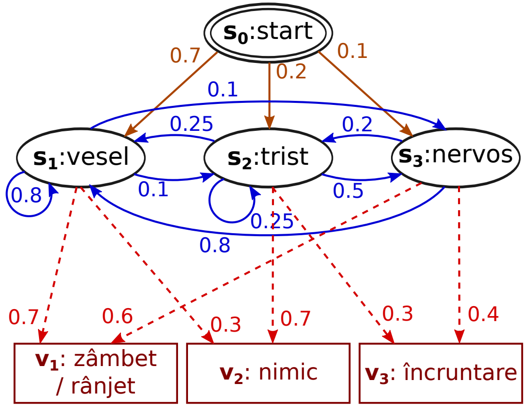
Exemplu: Urmărirea stărilor emoționale



vesel → vesel → vesel → trist → nervos

zâmbet, nimic, nimic, încruntare, zâmbet

nervos → trist → trist → nervos → vesel

O - secvența de observații

T - lungimea secvenței de observații

$$O = [o_1 o_2 \cdots o_T]$$

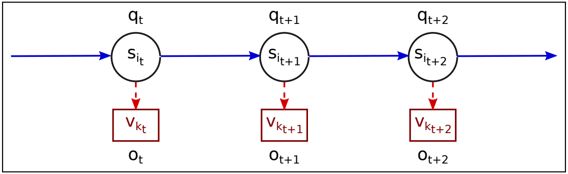
Modele Markov Ascunse

Definiție

Un Model Markov Ascuns este un tuplu $\langle S, V, A, B, P \rangle$:

- S - mulțimea stărilor
- V - mulțimea valorilor observabile
- A - matricea de tranziție
- B - matricea de emisie
- Π - matricea distribuției stării inițiale

- Notație: $\lambda = (A, B, \Pi)$ - parametrii modelului



Reformularea problemelor fundamentale ale MMA

Problema evaluării

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum calculăm probabilitatea $P(O|\lambda)$ ca secvența de observații să fi fost generată de acel model?

- Prin enumerarea tuturor secvențelor posibile de stări:

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q, \lambda) \cdot P(Q|\lambda) \tag{1}$$

(legea probabilității totale)

Reformularea problemelor fundamentale ale MMA

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q,\lambda) \cdot P(Q|\lambda) \tag{1}$$

Primul factor (independență condițională):

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(o_t|q_t,\lambda) = \prod_{t=1}^{T} b_{q_t}(o_t) = b_{q_1}(o_1) \cdot \ldots \cdot b_{q_T}(o_T) \tag{2}$$

Al doilea factor (presupunerea Markov):

$$P(Q|\lambda) = \pi_{q_1} \prod_{t=2}^{T} a_{q_{t-1},q_t} = \pi_{q_1} \cdot a_{q_1,q_2} \cdot a_{q_2,q_3} \cdot \ldots \cdot a_{q_{T-1},q_T} \tag{3}$$

$$P(O|\lambda) = \sum_{\text{all } Q} \left(\pi_{q_1} \cdot b_{q_1}(o_1) \cdot \prod_{t=2}^{T} b_{q_t}(o_t) a_{q_{t-1},q_t} \right) \tag{1}$$

Reformularea problemelor fundamentale ale MMA

Problema interpretării unei secvențe de observații

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum alegem o secvență corespunzătoare de stări $Q = [q_1 q_2 \cdots q_T]$ care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

- Cum alegem criteriul pentru *cea mai bună* secvență?
  - Secvența celor mai probabile stări (luate individual):

$$Q_{\text{best}} = [q_{1_{\text{best}}} \; q_{2_{\text{best}}} \; \ldots \; q_{T_{\text{best}}}], \quad q_{t_{\text{best}}} = \underset{s_i}{argmax} \; P(q_t = s_i | O, \lambda) \tag{4}$$

  - Cea mai bună *cale* (de dimensiune T)

$$Q_{\text{best}} = \underset{Q}{\operatorname{argmax}} \; P(Q|O, \lambda) = \underset{Q}{\operatorname{argmax}} \; P(Q, O|\lambda) \tag{5}$$

Algoritmul Forward-Backward

Problema Evaluării

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \dots o_T]$, care este probabilitatea ca secvența de observații să fi fost produsă de acel model?

$$P(O|\lambda) = ? \tag{7}$$

Rezolvare

$P(O|\lambda)$ se calculează *eficient* cu algoritmul Forward - Backward.

Calculul se face cu ajutorul unor variabile auxiliare α și β.

Variabilele α - Motivație

- Calcul conform legii probabilităților totale:

$$P(O|\lambda) = \sum_{all\ Q} P(O, Q|\lambda) = \sum_{all\ Q} P(O, |Q, \lambda) \cdot P(Q, \lambda)$$
$$= \sum_{\text{all } Q} \left(\pi_{q_1} \cdot b_{q_1}(o_1) \cdot \prod_{t=2}^{T} b_{q_t}(o_t) a_{q_{t-1},q_t} \right) \tag{1}$$

- Pentru secvențele de stări care încep cu o secvență comună $[q_1 q_2 \ldots q_Z]$, următorii factori sunt comuni:

$$\pi_{q_1} \cdot b_{q_1}(o_1) \cdot \prod_{z=2}^{Z} b_{q_z}(o_z) a_{q_{z-1},q_z}$$

  - Calculul de $(T - Z)^N$ ori ar fi redundant!

Variabilele α - Definiție

Definim variabilele α astfel:

$$\alpha_{t,i} = \underset{1 \le t \le T, 1 \le i \le N}{P(o_1, o_2, \ldots, o_t, q_t = s_i | \lambda)} \tag{8}$$

(probabiliatea de a fi observat primele t valori observate ajungând la momentul t în starea s_i, dați fiind parametrii λ)

Variabilele α - Definiție

Definim variabilele α astfel:

$$\alpha_{t,i} = P(o_1, o_2, \ldots, o_t, q_t = s_i|\lambda) \quad 1 \le t \le T, 1 \le i \le N \tag{8}$$

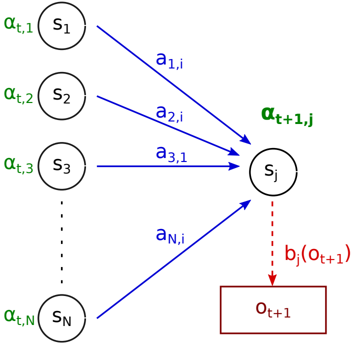
(probabiliatea de a fi observat primele t valori observate ajungând la momentul t în starea s_i, dați fiind parametrii λ)

Relația dintre $P(O|\lambda)$ și variabilele α este:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_{T,i} \tag{9}$$

(conform legii probabilităților totale)

Calculul variabilelor α



- Inițializare ($t = 1$)

$$\alpha_{1,i} = \pi_i b_i(o_1), \quad 1 \leq i \leq N$$

- Inducție ($t > 1$)

$$\alpha_{t+1,j} = \left[\sum_{i=1}^{N} \alpha_{t,j} a_{i,j} \right] b_j(o_{t+1}), \quad \substack{1 \leq t \leq T-1 \\ 1 \leq j \leq N}$$

- Probabilitatea secvenței observate:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_{T,i}$$

Variabilele β - Definiție

Definim variabilele β astfel:

$$\beta_{t,i} = P(o_{t+1} o_{t+2} \cdots o_T | q_t = s_i, \lambda) \tag{10}$$

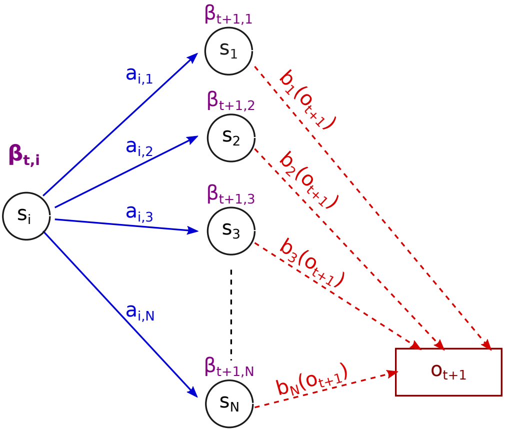
(probabiliatea de a fi observate valorile secvenței de la $t + 1$ la T, condiționată de aflarea la momentul t în starea s_i și de parametrii λ)

Relația dintre $P(O|\lambda)$ și variabilele β este:

$$P(O|\lambda) = \sum_{i=1}^{N} \beta_{1,i} \tag{11}$$

(conform legii probabilităților totale)

Calculul Variabilelor β



- Inițializare ($t = T$)

$$\beta_{T,i} = 1, \quad 1 \leq i \leq N \tag{12}$$

- Pas de inducție ($t < T$)

$$\beta_{t,i} = \sum_{j=1}^{N} a_{i,j} b_j(o_{t+1}) \beta_{t+1,j}, \quad t = T-1, T-2, \ldots, 1, 1 \leq i \leq N$$

Problema interpretării

Problema interpretării unei secvențe de observații

Date fiind un model $\lambda = (A, B, \Pi)$ și o secvență de observații $O = [o_1 o_2 \cdots o_T]$, cum alegem o secvență corespunzătoare de stări $Q_{\text{best}} = [q_{1_{\text{best}}} q_{2_{\text{best}}} \cdots q_{T_{\text{best}}}]$ care *să dea un înțeles* observațiilor? Cum *descoperim* partea ascunsă a modelului?

- Răspunsul depinde de criteriul cu care alegem *cea mai bună* secvență
  - secvența celor mai probabile stări $q_{t_{\text{best}}}$ (luate individual), date fiind modelul și secvența observată:
    $q_{t_{\text{best}}} = \underset{s_i}{\text{argmax}}\, P(q_t = s_i | O, \lambda)$
  - cea mai probabilă secvență de stări Q (per ansamblu), date fiind modelul și secvența observată:
    $Q_{\text{best}} = \underset{\text{all } Q}{\text{argmax}}\, P(Q | O, \lambda)$

Secvența celor mai probabile stări

Notație: $P(q_t = s_i | O, \lambda) = \gamma_{\mathbf{t,i}} = \frac{\alpha_{t,i}\beta_{t,i}}{\sum_{j=1}^{N} \alpha_{t,j}\beta_{t,j}}$

- Este un criteriu satisfăcător?

Secvența celor mai probabile stări

Notație: $P(q_t = s_i | O, \lambda) = \gamma_{\mathbf{t,i}} = \frac{\alpha_{t,i}\beta_{t,i}}{\sum_{j=1}^{N} \alpha_{t,j}\beta_{t,j}}$

- Este un criteriu satisfăcător?
- **NU!**
  Pot exista q_t și q_{t+1} astfel încât $a_{q_t,q_{t+1}} = 0$

Algoritmul Viterbi

- Criteriu care ia în considerare distribuțiile de probabilitate ale tranzițiilor între stări
- Cea mai bună cale: $Q_{\text{best}} = [q_{1_{\text{best}}} q_{2_{\text{best}}} \cdots q_{T_{\text{best}}}]$

$$Q_{\text{best}} = \underset{Q}{\text{argmax}}\ P(Q|O,\lambda) = \underset{Q}{\text{argmax}}\ P(Q,O|\lambda) \tag{6}$$

- **Algoritmul Viterbi** - programare dinamică [Vit67]
- Vom introduce variabilele δ.

Variabilele δ - intuiție

Întrebare

Dacă $Q = [q_1, q_2, \ldots, q_{t-1}, q_t]$ este cea mai bună secvență care explică $O = [o_1, o_2, \ldots o_{t-1}, o_t]$, atunci putem afirma că $Q[1 : t-1]$ este cea mai bună secvență de stări care explică $O[1 : t-1]$?

Variabilele δ - intuiție

Întrebare

Dacă $Q = [q_1, q_2, \ldots, q_{t-1}, q_t]$ este cea mai bună secvență care explică $O = [o_1, o_2, \ldots o_{t-1}, o_t]$, atunci putem afirma că $Q[1:t-1]$ este cea mai bună secvență de stări care explică $O[1:t-1]$?

- **NU!**

$$P(q_1 = s_{i_1}, q_2 = s_{i_2}, \ldots, q_{t-1} = s_{i_{t-1}}, q_t = s_{i_t} | O, \lambda) =$$
$$P(q_1, q_2, \ldots, q_{t-1} | O, \lambda) \cdot P(q_t = s_{i_t} | q_{t-1} = s_{i_{t-1}}, \lambda) \cdot P(o_t | q_t = s_{i_t}, \lambda)$$

Variabilele δ

- Vom numi variabile δ:

$$\delta_{t,i} = \max_{q_1,\dots,q_{t-1}} P([q_1 q_2 \dots q_{t-1} s_i], [o_1, o_2, \dots o_t] | \lambda) \qquad (26)$$

- $\delta_{\mathbf{t},\mathbf{i}}$ - cea mai mare probabilitate a unei secvențe de stări de lungime t care ajunge în s_i și explică primele t valori observate

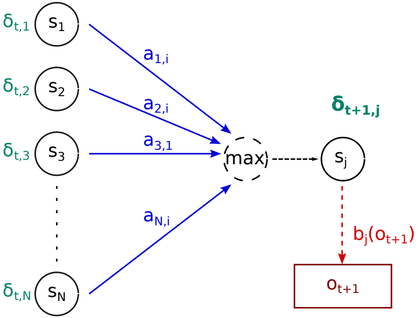
Variabilele δ

- Vom numi variabile δ:

$$\delta_{t,i} = \max_{q_1,\ldots,q_{t-1}} P([q_1 q_2 \ldots q_{t-1} s_i], [o_1, o_2, \ldots o_t] | \lambda) \quad (26)$$

- $\delta_{\mathbf{t,i}}$ - cea mai mare probabilitate a unei secvențe de stări de lungime t care ajunge în s_i și explică primele t valori observate
- relația dintre variabilele δ:

$$\delta_{t+1,j} = [\max_i \delta_{t,i} \cdot a_{i,j}] \cdot b_j(o_{t+1}) \quad (27)$$



Algoritmul Viterbi - Privire de ansamblu

Pași înainte

Calculăm cea mai mare probabilitate de a ajunge în starea s_i la momentul t pe baza celor mai mari probabilități de a fi ajuns în toate stările s_j la $t-1$.

Algoritmul Viterbi - Privire de ansamblu

Pași înainte

Calculăm cea mai mare probabilitate de a ajunge în starea s_i la momentul t pe baza celor mai mari probabilități de a fi ajuns în toate stările s_j la $t-1$.

La final ($t = T$)

Starea finală este acea stare s_i cu cea mai mare probabilitate de a ajunge la ea după observarea tuturor valorilor.

Algoritmul Viterbi - Privire de ansamblu

Pași înainte

Calculăm cea mai mare probabilitate de a ajunge în starea s_i la momentul t pe baza celor mai mari probabilități de a fi ajuns în toate stările s_j la $t-1$.

La final ($t = T$)

Starea finală este acea stare s_i cu cea mai mare probabilitate de a ajunge la ea după observarea tuturor valorilor.

Pași înapoi

Refacem *cea mai bună cale* alegând pentru fiecare moment t starea care a dus la probabilitatea maximă pentru starea de la $t+1$.

Algoritmul Viterbi (I)

1 Inițializare:

$$\begin{aligned} \delta_{1,i} &= \pi_i b_i(o_1), \quad 1 \le i \le N \\ \psi_{1,i} &= 0 \end{aligned} \tag{28}$$

2 Recursivitate:

$$\begin{aligned} \delta_{t,j} &= [\max_i \delta_{t-1,i} \cdot a_{i,j}] \cdot b_j(o_t) \quad {\scriptstyle 2 \le t \le T, 1 \le j \le N} \\ \psi_{t,i} &= \operatorname*{argmax}_i \delta_{t-1,i} \cdot a_{i,j} \quad {\scriptstyle 2 \le t \le T, 1 \le j \le N} \end{aligned} \tag{29}$$

Algoritmul Viterbi (II)

3 Terminare:

$$P(Q_{\text{best}}|O,\lambda) = \max_i \delta_{T,i}$$
$$q_{T_{\text{best}}} = \operatorname*{argmax}_i \delta_{T,i} \tag{30}$$

4 Backtracking:

$$q_{t_{\text{best}}} = \psi_{t+1}(q_{t+1_{\text{best}}}), \quad t=T-1,T-2,\cdots,1 \tag{31}$$

Probleme numerice

- Cine este $P(Q_{\text{best}}|O,\lambda)$?

$$
\begin{aligned}
P(Q_{\text{best}}|O,\lambda) &= \delta_{T,i_T} \\
P(Q_{\text{best}}|O,\lambda) &= \delta_{T-1,i_{T-1}} \cdot a_{i_{T-1},i_T} \cdot b_T(o_T) \\
P(Q_{\text{best}}|O,\lambda) &= \delta_{T-2,i_{T-2}} \cdot a_{i_{T-2},i_{T-1}} \cdot b_{T-1}(o_{T-1}) \cdot a_{i_{T-1},j_T} \cdot b_T(o_T) \\
P(Q_{\text{best}}|O,\lambda) &= \prod \ldots
\end{aligned}
$$

- Cum putem evita apropierea rapidă de zero?

Probleme numerice

- Cine este $P(Q_{\text{best}}|O,\lambda)$?

$$P(Q_{\text{best}}|O,\lambda) = \delta_{T,i_T}$$
$$P(Q_{\text{best}}|O,\lambda) = \delta_{T-1,i_{T-1}} \cdot a_{i_{T-1},i_T} \cdot b_T(o_T)$$
$$P(Q_{\text{best}}|O,\lambda) = \delta_{T-2,i_{T-2}} \cdot a_{i_{T-2},i_{T-1}} \cdot b_{T-1}(o_{T-1}) \cdot a_{i_{T-1},j_T} \cdot b_T(o_T)$$
$$P(Q_{\text{best}}|O,\lambda) = \prod \ldots$$

- Cum putem evita apropierea rapidă de zero?
- Calculăm $log(P)$

$$\log(P(Q_{\text{best}}|O,\lambda)) = \log(\delta_{T,i_T})$$
$$\log(P(Q_{\text{best}}|O,\lambda)) = \log(\delta_{T-1,i_{T-1}}) + \log(a_{i_{T-1},i_T}) + \log(b_T(o_T))$$
$$\log(P(Q_{\text{best}}|O,\lambda)) = \log(\delta_{T-2,i_{T-2}}) + \log(a_{i_{T-2},i_{T-1}}) + \log(b_{T-1}(o_{T-1})$$
$$\log(P(Q_{\text{best}}|O,\lambda)) = \sum \ldots$$

Probleme numerice - Rezolvare

- Notație:

$$\phi_{t,i} = \max_{q_1,\ldots,q_{t-1}} \log(P(q_1,\ldots,q_{t-1},q_t = s_i, o_1,\ldots,o_t|\lambda)) = \log(\delta_{t,i})$$

- Matricele $\log(\Pi)$, $\log(A)$ și $\log(B)$ pot fi precalculate.

Algoritmul Viterbi - $log(P)$ (I)

1 Inițializare:

$$\begin{aligned} \phi_{1,i} &= \log(\pi_i) + \log(b_i(o_1)), \quad 1 \le i \le N \\ \psi_{1,i} &= 0 \end{aligned} \tag{32}$$

2 Recursivitate:

$$\begin{aligned} \phi_{t,j} &= [\max_i \phi_{t-1,i} + log(a_{i,j})] + \log(b_j(o_t)) \quad {\scriptstyle 2 \le t \le T, 1 \le j \le N} \\ \psi_{t,i} &= \operatorname{argmax}_i \phi_{t-1,i} + \log(a_{i,j}) \quad {\scriptstyle 2 \le t \le T, 1 \le j \le N} \end{aligned} \tag{33}$$

Algoritmul Viterbi - *log*(*P*) (II)

3 Terminare:

$$\log(P(Q_{\text{best}}|O,\lambda)) = \max_i \phi_{T,i}$$
$$q_{T_{\text{best}}} = \operatorname*{argmax}_i \phi_{T,i} \tag{34}$$

4 Backtracking:

$$q_{t_{\text{best}}} = \psi_{t+1}(q_{t+1_{\text{best}}}), \quad {\scriptstyle t=T-1,T-2,\cdots,1} \tag{35}$$