Sistemas Inteligentes

Escuela Técnica Superior de Informática

Universitat Politècnica de València

Tema B2T6:
Algoritmo de Viterbi.
Estimación de modelos de Markov.
Re-estimación por Viterbi.

índice

∘ 1 *Algoritmo de Viterbi* ▷ 1

2 Aprendizaje: estimación de probabilidades en modelos de Markov ▷ 10

3 Inicialización de la re-estimación por Viterbi ▷ 19

Aproximación de Viterbi a $P(y \mid M)$

Dado un modelo oculto de Markov $M = (Q, \Sigma, \pi, A, B)$ con estado final F, y una cadena $y = y_1 \cdots y_m \in \Sigma^+$, la probabilidad de que M genere y es:

$$P(y \mid M) \;=\; \sum_{z \in Q^+} P(y, z) \;=\; \sum_{q_1, \ldots, q_m \in Q^+} P(y,\, q_1, \ldots, q_m)$$

Una aproximación a $P(y \mid M)$ es la llamada *aproximación de Viterbi*:

$$\tilde{P}(y \mid M) = \max_{q_1, \ldots, q_m \in Q^+} P(y,\, q_1, \ldots, q_m)$$

La correspondiente *secuencia de estados más probable* es:

$$\tilde{q} = (\tilde{q}_1, \ldots, \tilde{q}_m) = \operatorname*{argmax}_{q_1, \ldots, q_m \in Q^+} P(y,\, q_1, \ldots, q_m)$$

Algoritmo de Viterbi

Definimos $V(q,t)$ como la probabilidad máxima de que un modelo oculto de Markov M alcance el estado q en el instante t, emitiendo el prefijo $y_1 \dots y_t$:

$$V(q,t) = \max_{\substack{q_1,\dots,q_t \\ q_t=q}} P(y_1 \cdots y_t,\, q_1,\dots,q_t)$$

$V(q,t)$ puede calcularse recursivamente:

$$V(q,t) = \max_{\substack{q_1,\dots,q_t \\ q_t=q}} P(y_1 \cdots y_t,\, q_1,\dots,q_t)$$

$$= \max_{q' \in Q} \max_{\substack{q_1,\dots,q_{t-1} \\ q_{t-1}=q'}} P(y_1 \cdots y_{t-1},\, q_1,\dots,q_{t-1}) \cdot A_{q',q}\, B_{q,y_t}$$

$$= \max_{q' \in Q} V(q',t-1) \cdot A_{q',q}\, B_{q,y_t}$$

En general: $V(q,t) = \begin{cases} \pi_q\, B_{q,y_1} & \textit{\textbf{si}}\ t=1 \\ \max_{q' \in Q} V(q',t-1)\, A_{q',q}\, B_{q,y_t} & \textit{\textbf{si}}\ t>1 \end{cases}$

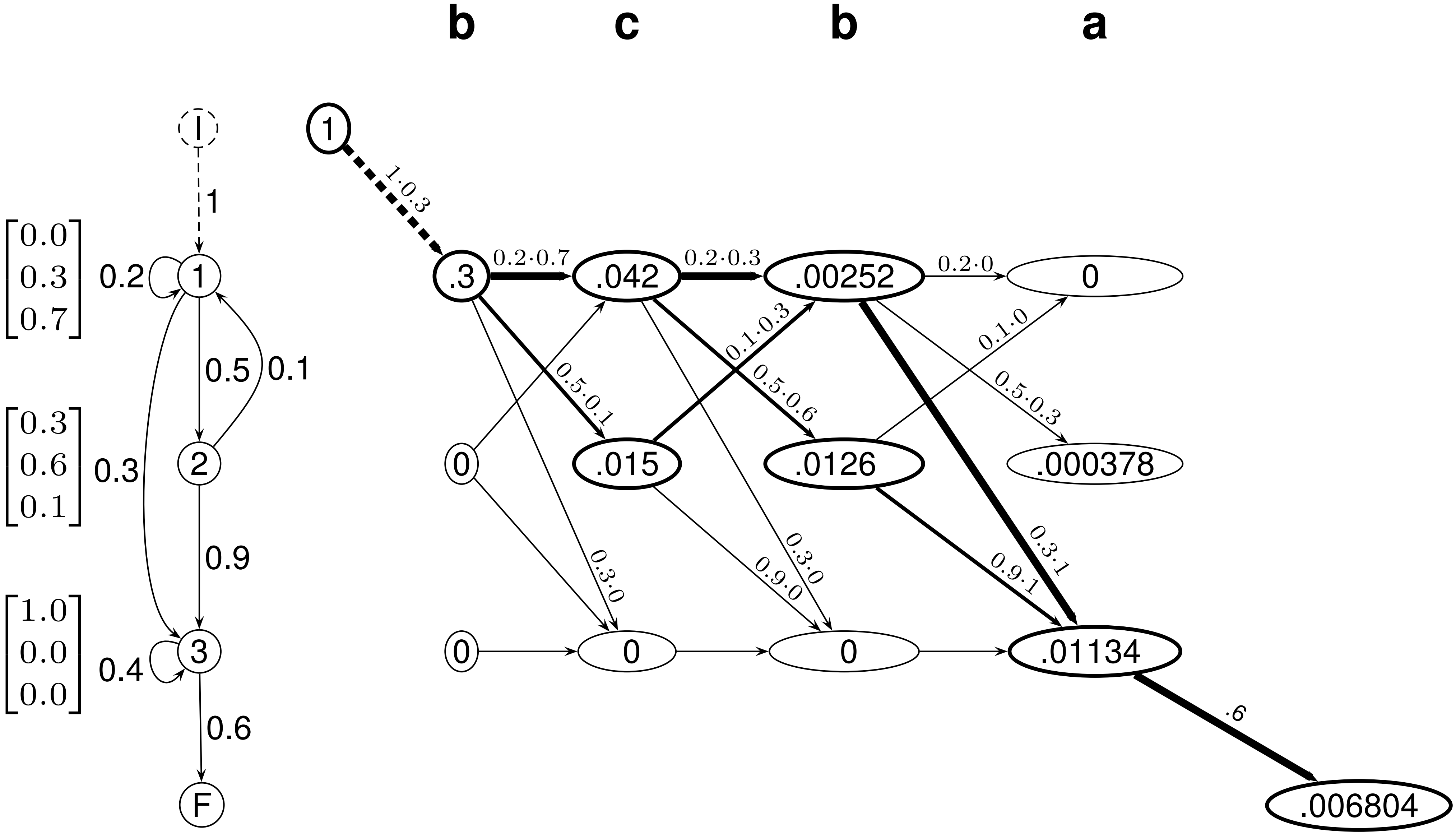
Algoritmo de Viterbi (cont.)

Aproximación de Viterbi a $P(y \mid M)$:

$$\tilde{P}(y \mid M) = \max_{q \in Q} V(q, |y|) \, A_{q,F}$$

- La función $V()$ puede representarse como una matriz: $V_{q,t} \equiv V(q, t)$.
- Esta matriz define un *grafo multietapa* denominado *trellis* y permite el *cálculo iterativo eficiente* de $V(q, |y|)$ *por Programación Dinámica*.
- La correspondiente secuencia óptima de estados, $\tilde{q}$, se calcula recorriendo el *trellis* hacia atrás.
- Complejidad temporal del algoritmo: $O(mb)$, donde m es la longitud de la cadena y b es el número de transiciones entre estados.

Algoritmo de Viterbi: ejemplo



Algoritmo de Viterbi: ejercicio

Sea M un modelo con:

$Q = \{1, 2, 3, F\}$

$\Sigma = \{a, b, c\}$

$\pi_1 = \pi_2 = \frac{1}{2}$, $\pi_3 = 0$

A	1	2	3	F
1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0
2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
3	0	0	$\frac{1}{2}$	$\frac{1}{2}$

B	a	b	c
1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
3	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

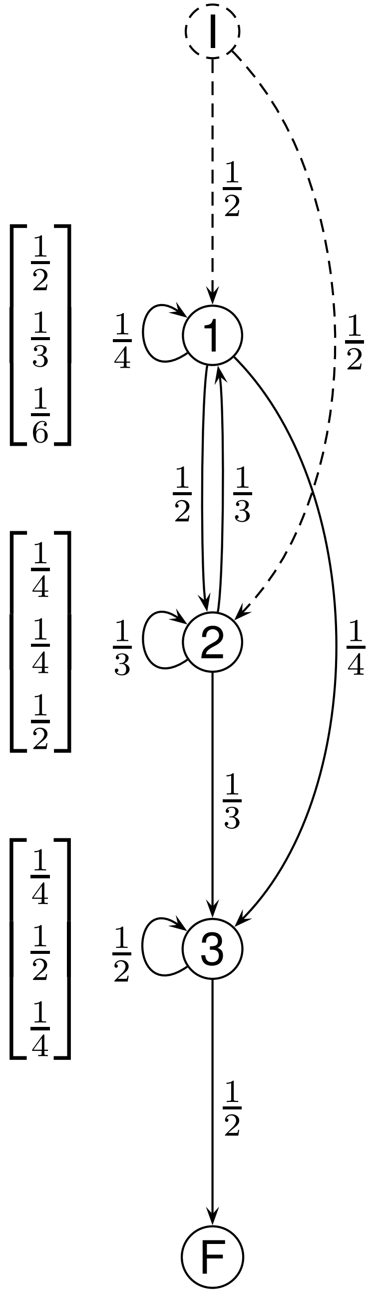
1. Halla el *trellis* para la cadena `abc`.
2. Obtén la secuencia óptima de estados asociada.

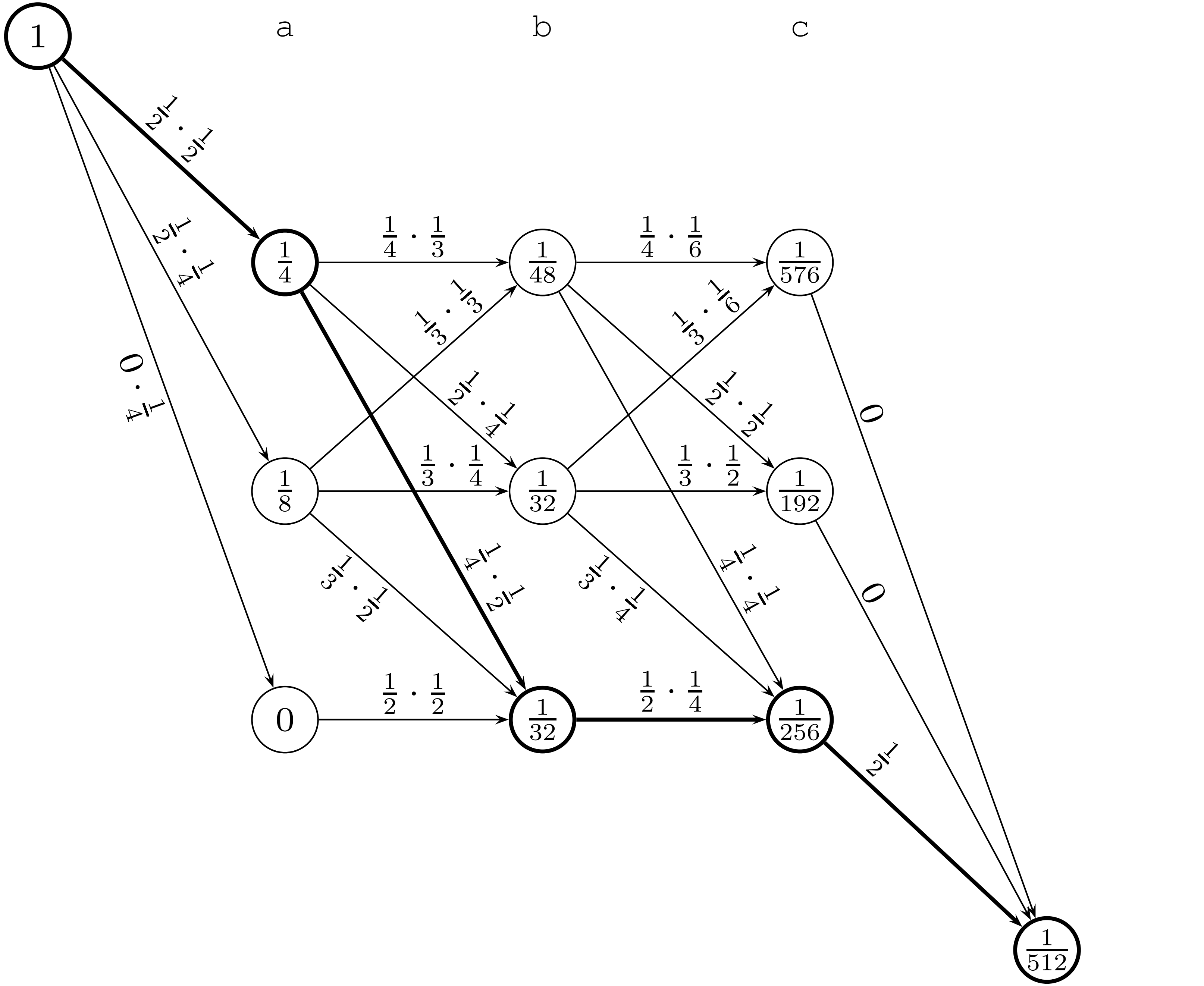
Ejercicio: resolución directa

V	a $t=1$	b $t=2$	c $t=3$	
1	$\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$	$\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{3}=\frac{1}{48}$ $\frac{1}{8}\cdot\frac{1}{3}\cdot\frac{1}{3}=\frac{1}{72}$	$\frac{1}{48}\cdot\frac{1}{4}\cdot\frac{1}{6}=\frac{1}{1152}$ $\frac{1}{32}\cdot\frac{1}{3}\cdot\frac{1}{6}=\frac{1}{576}$ $\frac{1}{32}\cdot 0\cdot\frac{1}{6}=0$	
2	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{32}$ $\frac{1}{8}\cdot\frac{1}{3}\cdot\frac{1}{4}=\frac{1}{96}$	$\frac{1}{48}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{192}$ $\frac{1}{32}\cdot\frac{1}{3}\cdot\frac{1}{2}=\frac{1}{192}$ $\frac{1}{32}\cdot 0\cdot\frac{1}{2}=0$	
3		$\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{2}=\frac{1}{32}$ $\frac{1}{8}\cdot\frac{1}{3}\cdot\frac{1}{2}=\frac{1}{48}$	$\frac{1}{48}\cdot\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{768}$ $\frac{1}{32}\cdot\frac{1}{3}\cdot\frac{1}{4}=\frac{1}{384}$ $\frac{1}{32}\cdot\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{256}$	
F				$\frac{1}{576}\cdot 0=0$ $\frac{1}{192}\cdot 0=0$ $\frac{1}{256}\cdot\frac{1}{2}=\frac{1}{512}$

$$\tilde{Q}=(1,3,3,F)$$

Ejercicio: resolución gráfica





Clasificación sintáctico-estadística mediante Viterbi

En la práctica, las probabilidades condicionales de las clases suelen aproximarse mediante Viterbi. Consideremos el ejercicio de la página B2T5.26:

$$
\begin{aligned}
\tilde{P}(y \mid M_A) &\\
&= \max(P(aab, q_1q_2q_3 = 001 \mid A),\\
&\qquad P(aab, q_1q_2q_3 = 011 \mid A),\\
&\qquad P(aab, q_1q_2q_3 = 111 \mid A))\\
&= \max(0.0544, 0.0086, 0.0008)\\
&= 0.0544
\end{aligned}
$$

$$
\begin{aligned}
\tilde{P}(y \mid M_B) &\\
&= \max(P(aab, q_1q_2q_3 = 001 \mid B),\\
&\qquad P(aab, q_1q_2q_3 = 011 \mid B))\\
&= \max(0.0680, 0.0108)\\
&= 0.0680
\end{aligned}
$$

$$\tilde{P}(A \mid y) = \frac{\tilde{P}(y \mid M_A)\, P(A)}{\sum_{c'} \tilde{P}(y \mid c')\, P(c')} = \frac{0.0544 \cdot 0.6}{0.0544 \cdot 0.6 + 0.0680 \cdot 0.4} = 0.5455$$

$$\tilde{P}(B \mid y) = 1 - \tilde{P}(A \mid y) = 0.4545$$

$\tilde{c}(y) = \operatorname*{argmax}_{c=A,B} \tilde{P}(c \mid y) = A$ resultado idéntico al de la página B2T5.27

índice

1 Algoritmo de Viterbi ▷ 1

◦ 2 *Aprendizaje: estimación de probabilidades en modelos de Markov* ▷ 10

3 Inicialización de la re-estimación por Viterbi ▷ 19

Estimación de probabilidades de un modelo de Markov

Problema básico:

Estimar las probabilidades de un modelo de Markov, M, mediante una secuencia de cadenas de entrenamiento $Y = \{y_1, \ldots, y_n\}$ extraidas independientemente de acuerdo con la ley de probabilidad $P(y|M)$.

Como las cadenas se han extraído independientemente:

$$P(Y|M) = \prod_{k=1}^{n} P(y_k|M)$$

El *estimador de máxima verosimilitud* de M es:

$$\hat{M} = \underset{M}{\text{argmax}} \prod_{k=1}^{n} P(y_k|M) \approx \underset{M}{\text{argmax}} \prod_{k=1}^{n} \tilde{P}(y_k|M)$$

Estimación mediante el algoritmo de Viterbi

Idea básica:

Analizar todas las cadenas Y, contabilizando las frecuencias de uso de las transiciones entre estados, de generación de símbolos en cada estado, etc. y normalizar adecuadamente.

Problema:
Cómo analizar las cadenas si no se conocen las probabilidades de del modelo?

Una posible solución:

1. Inicializar las probabilidades "adecuadamente"
2. Analizar cada cadena de Y mediante el algoritmo de Viterbi y obtener la secuencia de estados correspondiente
3. A partir de esta secuencia de estados, contabilizar las frecuencias requeridas.
4. Normalizar las frecuencias para obtener nuevas probabilidades del modelo
5. Repetir pasos 2-4 hasta convergencia.

Estimación mediante el algoritmo de Viterbi: ejemplo

Se dispone de tres cadenas de contorno de 4-direcciones que representan tres dígitos "*siete*" manuscritos[1].

A partir de estas cadenas se desea re-estimar las probabilidades de un modelo de Markov para estos dígitos. Utilizando el algoritmo de Viterbi se han obtenido las siguientes *secuencias óptimas de estados* para cada *cadena*:

Cadena:	**aaaaaddcdcdcdcdccbabaabababccccb**
Secuencia óptima de estados:	**11111222222222222333333333344444F**
Cadena:	**aaaaaddcdcdcdccdcbabababaabcccdcbb**
Secuencia óptima de estados:	**111112222222222223333333333444444 4F**
Cadena:	**aaaadcdcdcdcdcdcbababababccdccbaab**
Secuencia óptima de estados:	**1111222222222222333333333444444444F**

[1]Para mayor claridad se representan los trazos horizontales como *0=a, 2=c* y los verticales como *1=b, 3=d*.

Estimación mediante el algoritmo de Viterbi: ejemplo (cont.)

$$\pi_1 = 3/3 = 1 \quad \pi_2 = \pi_3 = \pi_4 = 0$$

A	1	2	3	4	F
1	$4+4+3$	$1+1+1$	0	0	0
2	0	$11+11+11$	$1+1+1$	0	0
3	0	0	$9+9+8$	$1+1+1$	0
4	0	0	0	$4+6+8$	$1+1+1$

$\Rightarrow$

A	1	2	3	4	F
1	$\frac{11}{14}$	$\frac{3}{14}$	0	0	0
2	0	$\frac{33}{36}$	$\frac{3}{36}$	0	0
3	0	0	$\frac{26}{29}$	$\frac{3}{29}$	0
4	0	0	0	$\frac{18}{21}$	$\frac{3}{21}$

B	a	b	c	d
1	$5+5+4$	0	0	0
2	0	0	$6+6+6$	$6+6+6$
3	$5+5+4$	$5+5+5$	0	0
4	$0+0+2$	$1+2+2$	$4+4+4$	$0+1+1$

$\Rightarrow$

B	a	b	c	d
1	$\frac{14}{14}$	0	0	0
2	0	0	$\frac{18}{36}$	$\frac{18}{36}$
3	$\frac{14}{29}$	$\frac{15}{29}$	0	0
4	$\frac{2}{21}$	$\frac{5}{21}$	$\frac{12}{21}$	$\frac{2}{21}$

Algoritmo de reestimación por Viterbi

Input: $M^0 = (Q^0, \Sigma^0, \pi^0, A^0, B^0)$ /* Modelo inicial */

$Y = \{y_1, \dots, y_n\}$ /* cadenas de entrenamiento */

Output: $M = (Q, \Sigma, \pi, A, B)$ /* Modelo optimizado */

$M = M^0$

repeat $M' = M;\ \pi = 0;\ A = 0; B = 0$

  for $k = 1$ **to** n **do**

    $m = |y_k|$ /* secuencia de estados más probable para y_k, */

    $\tilde{q}_1, \dots, \tilde{q}_m = \text{argmax}_{q_1,\dots,q_m} P(y_k\,, q_1, \dots, q_m \mid M')$ /* por Viterbi */

    $\pi_{\tilde{q}_1}$++; $B_{\tilde{q}_1, y_{k,1}}$++ /* actualización de contadores */

    for $t = 2$ **to** m **do** $A_{\tilde{q}_{t-1}, \tilde{q}_t}$++; $B_{\tilde{q}_t, y_{k,t}}$++ **done**; $A_{\tilde{q}_m, F}$ ++

  done

  $s = \sum_{q \in Q} \pi_q$

  forall $q \in Q$ **do** /* normalización de contadores */

    $\pi_q = \pi_q / s$

    $a = \sum_{q' \in Q} A_{q,q'}$; **forall** $q' \in Q$ **do** $A_{q,q'} = A_{q,q'}/a$

    $b = \sum_{\sigma \in \Sigma} B_{q,\sigma}$; **forall** $\sigma \in \Sigma$ **do** $B_{q,\sigma} = B_{q,\sigma}/b$

  done

until $M = M'$

Algoritmo mediante el algoritmo de Viterbi: ejercicio

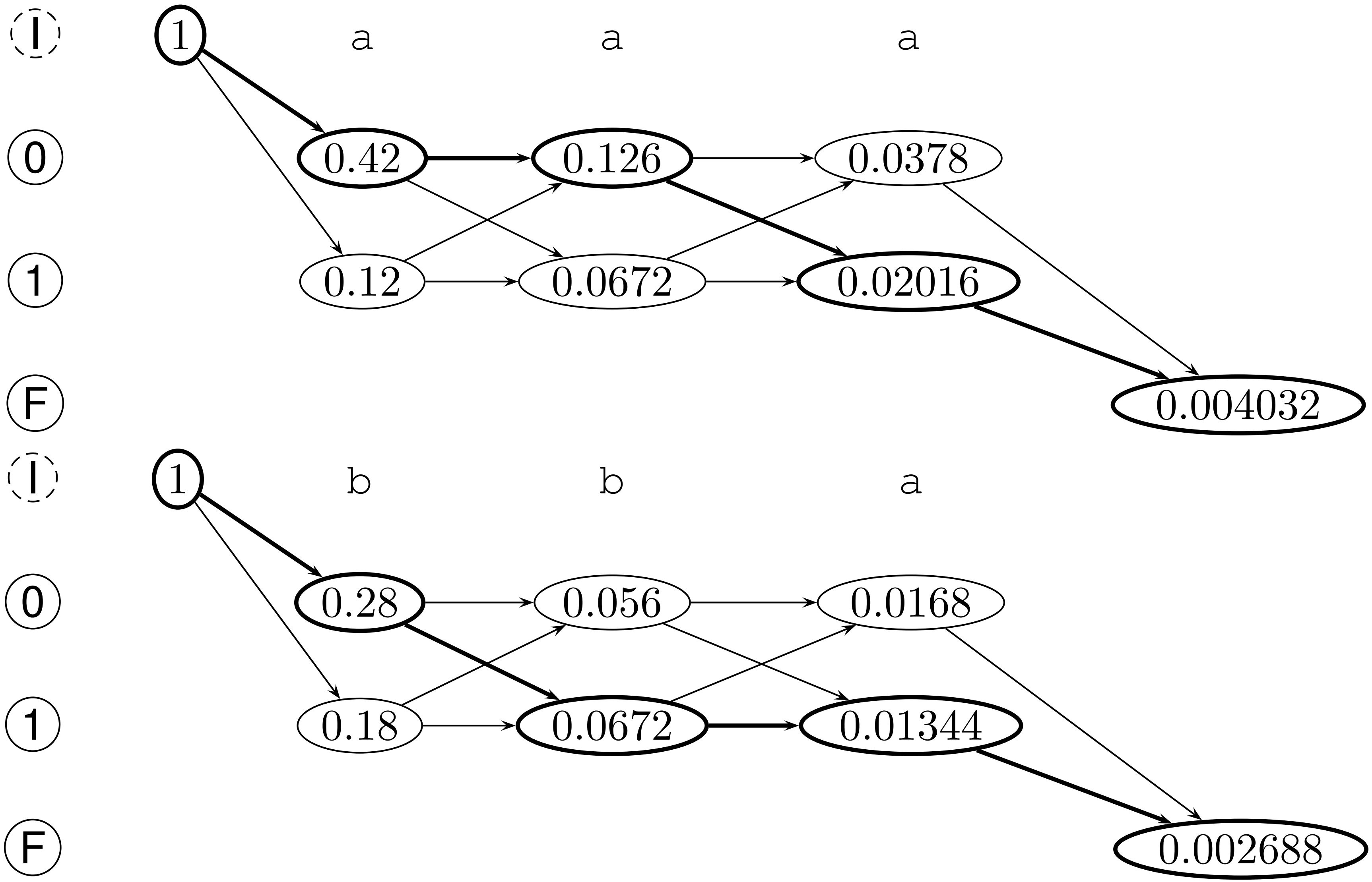
Sea M un modelo de Markov de conjunto de estados $Q = \{0, 1, F\}$; alfabeto $\Sigma = \{a, b\}$; probabilidades iniciales $\pi_0(0) = 0.7, \pi_0(1) = 0.3$; y probabilidades de transición entre estados y de emisión de símbolos:

A	0	1	F
0	0.5	0.4	0.1
1	0.3	0.5	0.2

B	a	b
0	0.6	0.4
1	0.4	0.6

Reestima los parámetros de M mediante *una* iteración de reestimación por Viterbi, a partir de las cadenas de entrenamiento "`a a a`" y "`b b a`".

Ejercicio: secuencias de estados mas probables



Los pares *cadena-secuencia óptima de estados* obtenidos son:

a a a	b b a
0 0 1 F	0 1 1 F

Ejercicio: parámetros reestimados

$$\hat{\pi}_0(0) = \frac{2}{2} = 1$$

$$\hat{\pi}_0(1) = \frac{0}{2} = 0$$

A	0	1	F
0	$\frac{1}{3}$	$\frac{2}{3}$	0
1	0	$\frac{1}{3}$	$\frac{2}{3}$

B	a	b
0	$\frac{2}{3}$	$\frac{1}{3}$
1	$\frac{2}{3}$	$\frac{1}{3}$

índice

1 Algoritmo de Viterbi ▷ 1

2 Aprendizaje: estimación de probabilidades en modelos de Markov ▷ 10

○ 3 *Inicialización de la re-estimación por Viterbi* ▷ 19

Inicialización de la re-estimación por Viterbi

Una idea elemental: Inicializar todas las probabilidades según distribuciones equiprobables.

Problema: Suele producir problemas de convergencia o convergencia a máximos locales inadecuados.

Una idea útil para modelos lineales:

- Segmentar cada cadena de Y en tantos segmentos de (aproximadamente) la misma longitud como estados tenga el modelo de Markov.
- Asignar los símbolos de cada segmento a su correspondiente estado
- Contabilizar las frecuencias de generación y transición
- Normalizar las frecuencias para obtener probabilidades iniciales requeridas

Inicialización por segmentación lineal: Ejemplo

Obtener un modelo de Markov de $N = 3$ estados mediante segmentación lineal a partir de las cadenas

$$y_1 = \text{aabbbcc} \qquad y_2 = \text{aaabbccc}$$

$$Q = \{1, 2, 3, F\} \qquad \Sigma = \{a, b, c\}$$

$$q = \left\lfloor \frac{t \cdot N}{\mid y \mid +1} \right\rfloor + 1 : \quad \begin{array}{ll} \text{aabbbcc} & \text{aaabbccc} \\ 1122233 & 11222333 \end{array}$$

$$\pi_1 = \frac{2}{2}, \quad \pi_2 = \pi_3 = 0$$

A	1	2	3	F
1	$\frac{2}{4}$	$\frac{2}{4}$	0	0
2	0	$\frac{4}{6}$	$\frac{2}{6}$	0
3	0	0	$\frac{3}{5}$	$\frac{2}{5}$

B	a	b	c
1	$\frac{4}{4}$	0	0
2	$\frac{1}{6}$	$\frac{5}{6}$	0
3	0	0	$\frac{5}{5}$

Inicialización por segmentación lineal

Input: $Y = \{y_1, \ldots, y_n\}, N$ /* cadenas de entrenamiento, número de estados */
Output: $M = (Q, \Sigma, \pi, A, B)$ /* modelo */

$Q = \{1, 2, \ldots, N, F\}; \Sigma = \{y \in y_k \in Y\}$ /* estados y símbolos */
$\pi = 0; A = 0; B = 0$ /* inicialización de contadores */

for $k = 1$ **to** n **do** /* actualización de contadores por */
$q = 1; \pi_q$++; $B_{q,y_{k,1}}$++ /* alineamiento lineal de y_k con los estados */
for $t = 2$ **to** $|y_k|$ **do** $q' = q; q = \left\lfloor \frac{t}{|y_k|+1} N \right\rfloor + 1; A_{q',q}$++; $B_{q,y_{k,t}}$++ **done**
$A_{q,F}$ ++
done

$s = \sum_{q \in Q} \pi_q$
forall $q \in Q$ **do** /* normalización de contadores */
$\pi_q = \pi_q \,/\, s$
$a = \sum_{q' \in Q} A_{q,q'}$; **forall** $q' \in Q$ **do** $A_{q,q'} = A_{q,q'} \,/\, a$
$b = \sum_{\sigma \in \Sigma} B_{q,\sigma}$; **forall** $\sigma \in \Sigma$ **do** $B_{q,\sigma} = B_{q,\sigma} \,/\, b$
done