Metoda HMM:
Ukryty Model Markowa

Przykład

- Jak oszacować średnią temperaturę w przeszłym okresie?
- Na podstawie pierścieni drzew
- Macierz A: p-wo zmian temperatur z roku na rok
- Macierz B: wpływ temp. na grubość pierścienia
- Stan początkowy: [0.6, 0.4]
- **Problem 1:**
  Dany jest ciąg obserwacji
  O = (S,M,S,L)
  Jaki był najbardziej prawdopodobny ciąg temperatur w tych latach?

$$\begin{array}{c} \\ H \\ C \end{array} \begin{array}{c} \begin{array}{cc} H & C \end{array} \\ \left[\begin{array}{cc} 0.7 & 0.3 \\ 0.4 & 0.6 \end{array}\right] \end{array}$$

$$\begin{array}{c} \\ H \\ C \end{array} \begin{array}{c} \begin{array}{ccc} S & M & L \end{array} \\ \left[\begin{array}{ccc} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{array}\right] \end{array}$$

$$\pi = \left[\begin{array}{cc} 0.6 & 0.4 \end{array}\right].$$

Notacje

$$
\begin{aligned}
T &= \text{the length of the observation sequence} \\
N &= \text{the number of states in the model} \\
M &= \text{the number of observation symbols} \\
Q &= \{q_0, q_1, \ldots, q_{N-1}\} = \text{the states of the Markov process} \\
V &= \{0, 1, \ldots, M-1\} = \text{set of possible observations} \\
A &= \text{the state transition probabilities} \\
B &= \text{the observation probability matrix} \\
\pi &= \text{the initial state distribution} \\
\mathcal{O} &= (\mathcal{O}_0, \mathcal{O}_1, \ldots, \mathcal{O}_{T-1}) = \text{observation sequence.}
\end{aligned}
$$

$$\mathcal{O}_i \in V \text{ for } i = 0, 1, \ldots, T-1$$

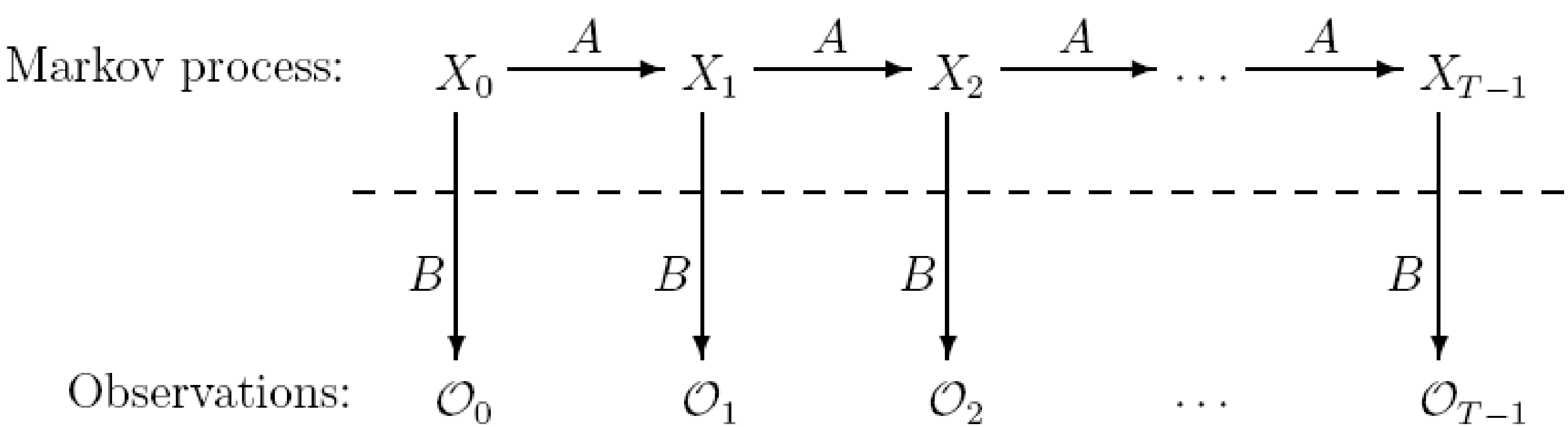
HMM to trójka: $\lambda = (A, B, \pi)$.

The matrix $A = \{a_{ij}\}$ is $N \times N$ with

$$a_{ij} = P(\text{state } q_j \text{ at } t+1 \,|\, \text{state } q_i \text{ at } t)$$

$B = \{b_j(k)\}$ is an $N \times M$ with

$$b_j(k) = P(\text{observation } k \text{ at } t \,|\, \text{state } q_j \text{ at } t).$$



Przykład obliczenia w HMM

- Niech $X = (x_0, x_1, x_2, x_3)$

$$O = (O_0, O_1, O_2, O_3)$$

$$P(X) = \pi_{x_0} b_{x_0}(\mathcal{O}_0) a_{x_0,x_1} b_{x_1}(\mathcal{O}_1) a_{x_1,x_2} b_{x_2}(\mathcal{O}_2) a_{x_2,x_3} b_{x_3}(\mathcal{O}_3).$$

- Np.

$$P(HHCC) = 0.6(0.1)(0.7)(0.4)(0.3)(0.7)(0.6)(0.1) = 0.000212.$$

3 problemy (Rabiner, 1989)

- **Problem 1: Oszacowanie**
  - Dany jest model $\lambda = (A, B, \pi)$ i ciąg obserwacji O;
  - Znaleźć $P(\mathrm{O} \mid \lambda)$
- **Problem 2: Ciąg stanów**
  - Dany jest model $\lambda = (A, B, \pi)$ i ciąg obserwacji O;
  - Znaleźć Q^*: $P(Q^* \mid O, \lambda) = \max_Q P(Q \mid O, \lambda)$
- **Problem 3: Uczenie się**
  - Dany jest ciąg obserwacji O, wymiary M,N;
  - Znaleźć model $\lambda = (A, B, \pi)$ maksymalizując $P(\mathrm{O} \mid \lambda)$

Problem1: Oszacowanie
Metoda 1

$$\begin{aligned} P(\mathcal{O}\,|\,\lambda) &= \sum_X P(\mathcal{O}, X\,|\,\lambda) \\ &= \sum_X P(\mathcal{O}\,|\,X, \lambda) P(X\,|\,\lambda) \\ &= \sum_X \pi_{x_0} b_{x_0}(\mathcal{O}_0) a_{x_0,x_1} b_{x_1}(\mathcal{O}_1) \cdots a_{x_{T-2},x_{T-1}} b_{x_{T-1}}(\mathcal{O}_{T-1}). \end{aligned}$$

- Czasochłonna metoda: $2TN^T$ mnożeń

Problem1: Oszacowanie

Metoda 2: „alpha-pass”

- Dla i = 0, …,T-1 oraz t = 0,…, N-1 definiujemy:

$$\alpha_t(i) = P(\mathcal{O}_0, \mathcal{O}_1, \ldots, \mathcal{O}_t, x_t = q_i \,|\, \lambda). \tag{8}$$

- Liczymy indukcyjnie:

1. Let $\alpha_0(i) = \pi_i b_i(\mathcal{O}_0)$, for $i = 0, 1, \ldots, N-1$

2. For $t = 1, 2, \ldots, T-1$ and $i = 0, 1, \ldots, N-1$, compute

$$\alpha_t(i) = \left[\sum_{j=0}^{N-1} \alpha_{t-1}(j) a_{ji}\right] b_i(\mathcal{O}_t)$$

3. Then from (8) it is clear that

$$P(\mathcal{O} \,|\, \lambda) = \sum_{i=0}^{N-1} \alpha_{T-1}(i).$$

Problem2: Ciąg stanów „beta-pass”

- Dla i = 0, …,T-1 oraz t = 0,…, N-1 definiujemy:

$$\beta_t(i) = P(\mathcal{O}_{t+1}, \mathcal{O}_{t+2}, \ldots, \mathcal{O}_{T-1} \,|\, x_t = q_i, \lambda).$$

- Liczymy indukcyjnie:

1. Let $\beta_{T-1}(i) = 1$, for $i = 0, 1, \ldots, N-1$.
2. For $t = T-2, T-1, \ldots, 0$ and $i = 0, 1, \ldots, N-1$ compute

$$\beta_t(i) = \sum_{j=0}^{N-1} a_{ij} b_j(\mathcal{O}_{t+1}) \beta_{t+1}(j).$$

For $t = 0, 1, \ldots, T-2$ and $i = 0, 1, \ldots, N-1$, define

$$\gamma_t(i) = P(x_t = q_i \,|\, \mathcal{O}, \lambda). = \frac{\alpha_t(i)\beta_t(i)}{P(\mathcal{O} \,|\, \lambda)}$$

Znaleźć ciąg stanów

- P: Czy wystarczy wziąć

$$q_t^* = \arg\max_i \gamma_t(i)$$

jako najlepszy ciąg?

- Odp.: **NIE** (przykład)

$\gamma_t(i)$	0	1	2	3
$P(H)$	0.188170	0.519432	0.228878	0.803979
$P(C)$	0.811830	0.480568	0.771122	0.196021

state	probability	normalized probability
$HHHH$	.000412	.042743
$HHHC$	.000035	.003664
$HHCH$	.000706	.073274
$HHCC$	.000212	.021982
$HCHH$	.000050	.005234
$HCHC$	.000004	.000449
$HCCH$	.000302	.031403
$HCCC$	.000091	.009421
$CHHH$	.001098	.113982
$CHHC$	.000094	.009770
$CHCH$	.001882	.195398
$CHCC$	.000564	.058619
$CCHH$	.000470	.048849
$CCHC$	.000040	.004187
$CCCH$	.002822	.293096
$CCCC$	.000847	.087929

Problem2: ciąg stanów
Algorytm Viterbi'ego

$$\delta_t(i) \equiv \max_{q1q2\cdots qt-1} p(q_1 q_2 \cdots q_{t-1}, q_t = S_i, O_1 \cdots O_t \mid \lambda)$$

- Initialization:

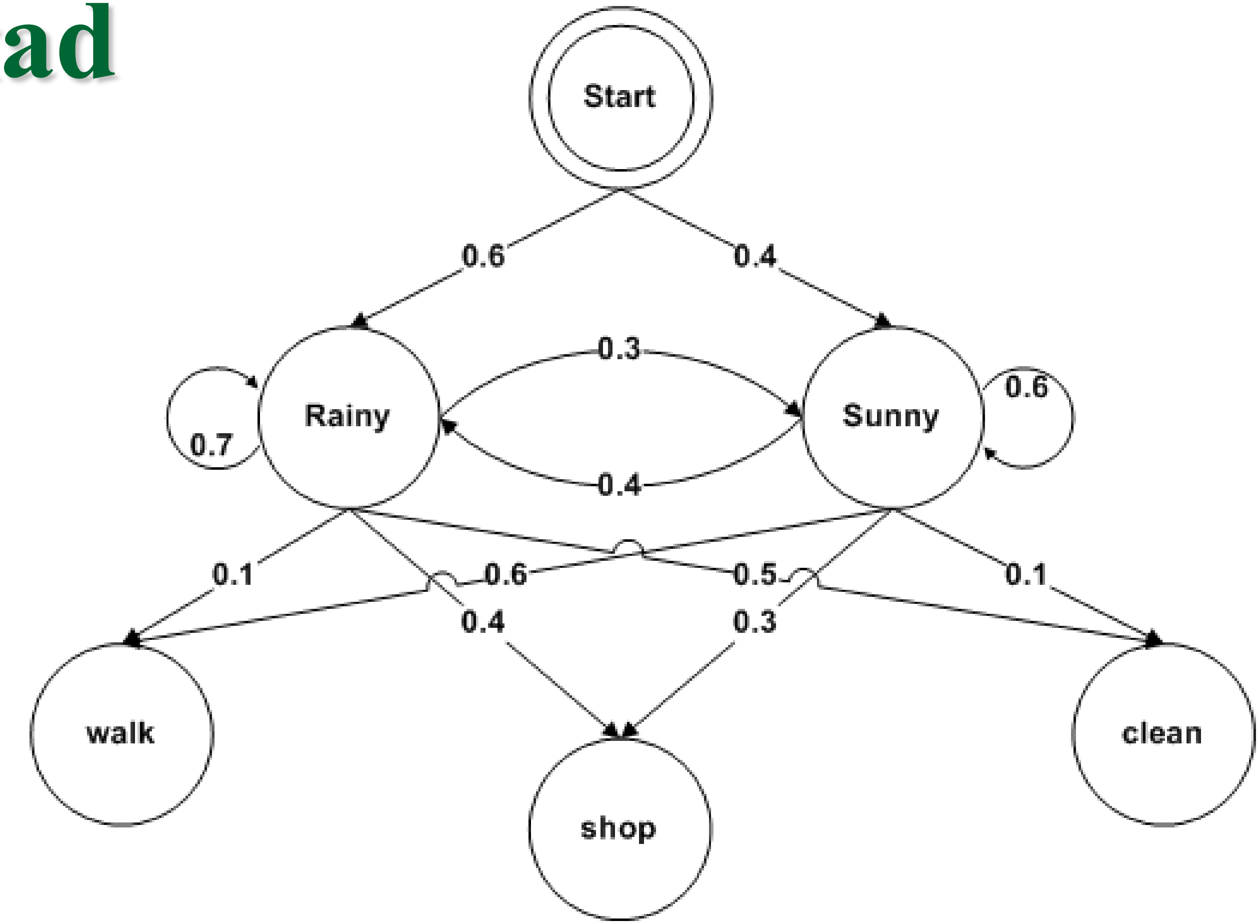
  $\delta_1(i) = \pi_i b_i(O_1),\ \psi_1(i) = 0$
- Recursion:

  $\delta_t(j) = \max_i \delta_{t-1}(i) a_{ij} b_j(O_t),\ \psi_t(j) = \mathrm{argmax}_i\ \delta_{t-1}(i) a_{ij}$
- Termination:

  $p^* = \max_i \delta_T(i),\ q_T^* = \mathrm{argmax}_i\ \delta_T(i)$
- Path backtracking:

  $q_t^* = \psi_{t+1}(q_{t+1}^*),\ t = T-1, T-2, \ldots, 1$

Przykład



Ciąg obserwacji: ['walk', 'shop', 'clean']
Ścieżka Viterbiego: ['Sunny', 'Rainy', 'Rainy', 'Rainy']

Problem 3: „uczenie się"
Algorytm Baum-Welch

- Definiujemy

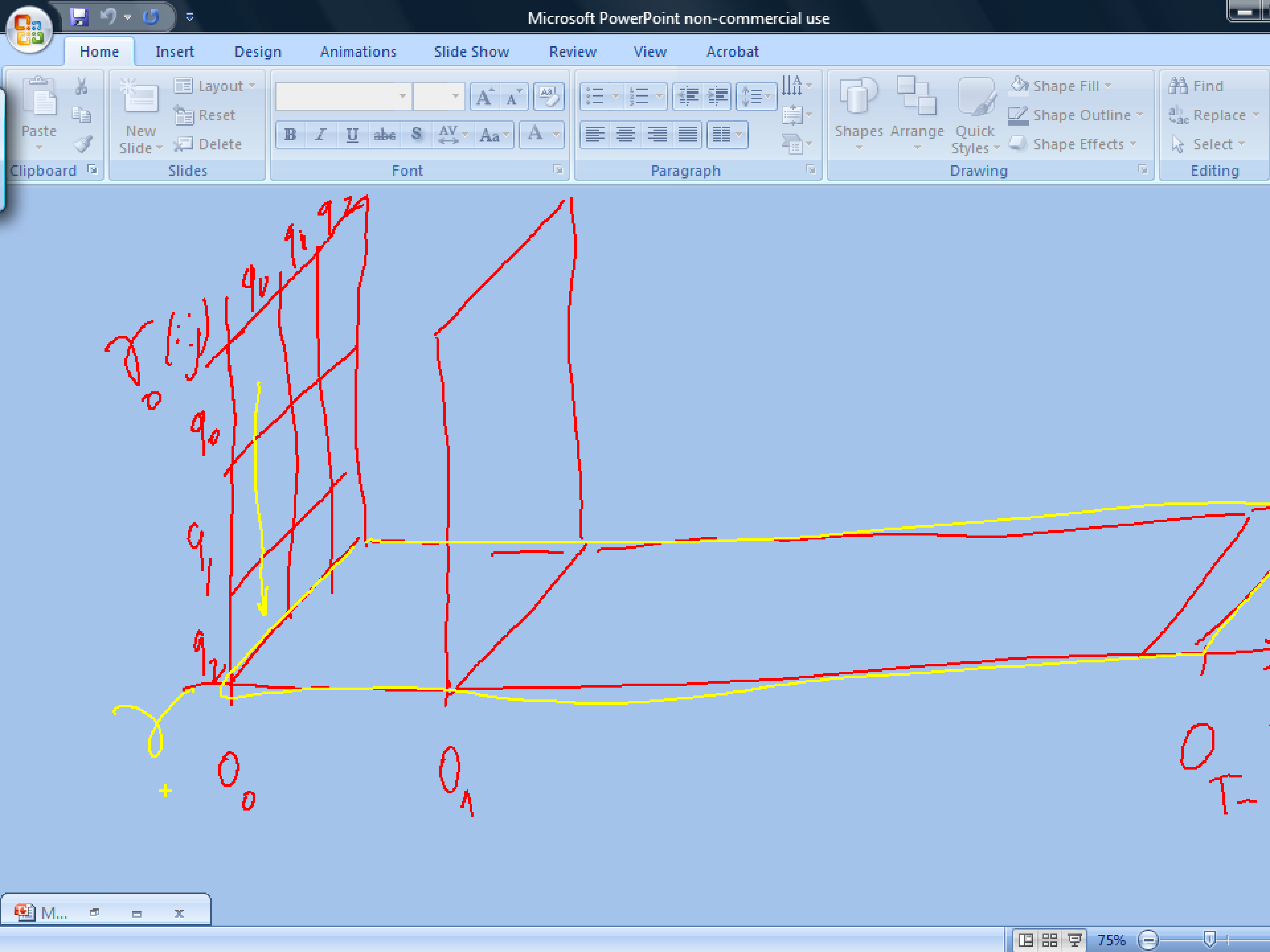
$$\gamma_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(\mathcal{O}_{t+1}) \beta_{t+1}(j)}{P(\mathcal{O} \mid \lambda)} \equiv P(q_t = S_i, q_{t+1} = S_j \mid O, \lambda)$$

- Wówczas

$$\gamma_t(i) = \sum_{j=0}^{N-1} \gamma_t(i,j).$$

- **Algorytm EM (Baum-Welch)**

1. Initialize, $\lambda = (A, B, \pi)$.
2. Compute $\alpha_t(i)$, $\beta_t(i)$, $\gamma_t(i,j)$ and $\gamma_t(i)$.
3. Re-estimate the model $\lambda = (A, B, \pi)$.
4. If $P(\mathcal{O} \mid \lambda)$ increases, goto 2.


Microsoft PowerPoint non-commercial use
Home
Insert
Design
Animations
Slide Show
Review
View
Acrobat
Paste
Clipboard
Layout
Reset
New Slide
Delete
Slides
Font
Paragraph
Shapes
Arrange
Quick Styles
Shape Fill
Shape Outline
Shape Effects
Drawing
Find
Replace
Select
Editing
75%

Algorytm Baum-Welch
- Re-estymacja modelu

For $i = 0, 1, \ldots, N-1$, let

$$\pi_i = \gamma_0(i)$$

For $i = 0, 1, \ldots, N-1$ and $j = 0, 1, \ldots, N-1$, compute

$$a_{ij} = \sum_{t=0}^{T-2} \gamma_t(i,j) \Bigg/ \sum_{t=0}^{T-2} \gamma_t(i)$$

For $j = 0, 1, \ldots, N-1$ and $k = 0, 1, \ldots, M-1$, compute

$$b_j(k) = \sum_{\substack{t \in \{0,1,\ldots,T-2\} \\ \mathcal{O}_t = k}} \gamma_t(j) \Bigg/ \sum_{t=0}^{T-2} \gamma_t(j)$$

Przykład (Cave & Neuwirth)

- HMM:
  - 2 stany
  - 27 symboli (26 i spacja)
  - Obserwacje: 50000 pierwszych liter pewnego tekstu z korpusu Browna.

$$\pi = \begin{bmatrix} 0.51316 & 0.48684 \end{bmatrix}$$

$$A = \begin{bmatrix} 0.47468 & 0.52532 \\ 0.51656 & 0.48344 \end{bmatrix}$$

- Experyment:
  - Początkowo: [1/2,1/2], …
  - 100 iteracji

$$\pi = \begin{bmatrix} 0.00000 & 1.00000 \end{bmatrix}$$

$$A = \begin{bmatrix} 0.25596 & 0.74404 \\ 0.71571 & 0.28429 \end{bmatrix}$$

	Initial		Final	
a	0.03735	0.03909	0.13845	0.00075
b	0.03408	0.03537	0.00000	0.02311
c	0.03455	0.03537	0.00062	0.05614
d	0.03828	0.03909	0.00000	0.06937
e	0.03782	0.03583	0.21404	0.00000
f	0.03922	0.03630	0.00000	0.03559
g	0.03688	0.04048	0.00081	0.02724
h	0.03408	0.03537	0.00066	0.07278
i	0.03875	0.03816	0.12275	0.00000
j	0.04062	0.03909	0.00000	0.00365
k	0.03735	0.03490	0.00182	0.00703
l	0.03968	0.03723	0.00049	0.07231
m	0.03548	0.03537	0.00000	0.03889
n	0.03735	0.03909	0.00000	0.11461
o	0.04062	0.03397	0.13156	0.00000
p	0.03595	0.03397	0.00040	0.03674
q	0.03641	0.03816	0.00000	0.00153
r	0.03408	0.03676	0.00000	0.10225
s	0.04062	0.04048	0.00000	0.11042
t	0.03548	0.03443	0.01102	0.14392
u	0.03922	0.03537	0.04508	0.00000
v	0.04062	0.03955	0.00000	0.01621
w	0.03455	0.03816	0.00000	0.02303
x	0.03595	0.03723	0.00000	0.00447
y	0.03408	0.03769	0.00019	0.02587
z	0.03408	0.03955	0.00000	0.00110
space	0.03688	0.03397	0.33211	0.01298

Eksperyment z 12 stanami

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	*T*	*U*	*V*	*W*	*X*	*Y*	*Z*	*SP*
V	*				*				*						*						*						
SP																											*
C						*	*			*	*																
LL											*													*			
FL	*									*					*		*										
VF																											
VP										*							*										
CF																								*			
1				*				*					*	*				*	*						*		
2					*				*						*						*						
3		*	*	*					*			*		*		*		*	*	*						*	
4								*				*			*			*		*	*						
5	*				*				*						*						*						
6		*	*	*		*	*	*		*		*	*	*		*	*	*	*			*	*		*	*	
7			*	*			*				*	*	*	*		*		*	*	*		*			*		
8					*							*								*							
9						*					*	*	*	*		*		*	*		*		*	*			
10	*				*				*						*												
11																											*
12			*			*	*									*			*				*				

Table 2: Cave and Neuwirth's 12 states result interpretation

Oznaczenia

V	Vowel
SP	Space
C	Consonant
FL	First Letter
LL	Last Letter
VF	Vowel Follower
VP	Vowel Proceeder
CP	Consonant Follower.