Hidden Markov Models

według **Bonnie Dorr** **Christof Monz**

Hidden Markov Model (HMM)

- HMM pozwala określić prawdopodobieństwo nieobserwowanej bezpośrednio sekwencji zdarzeń

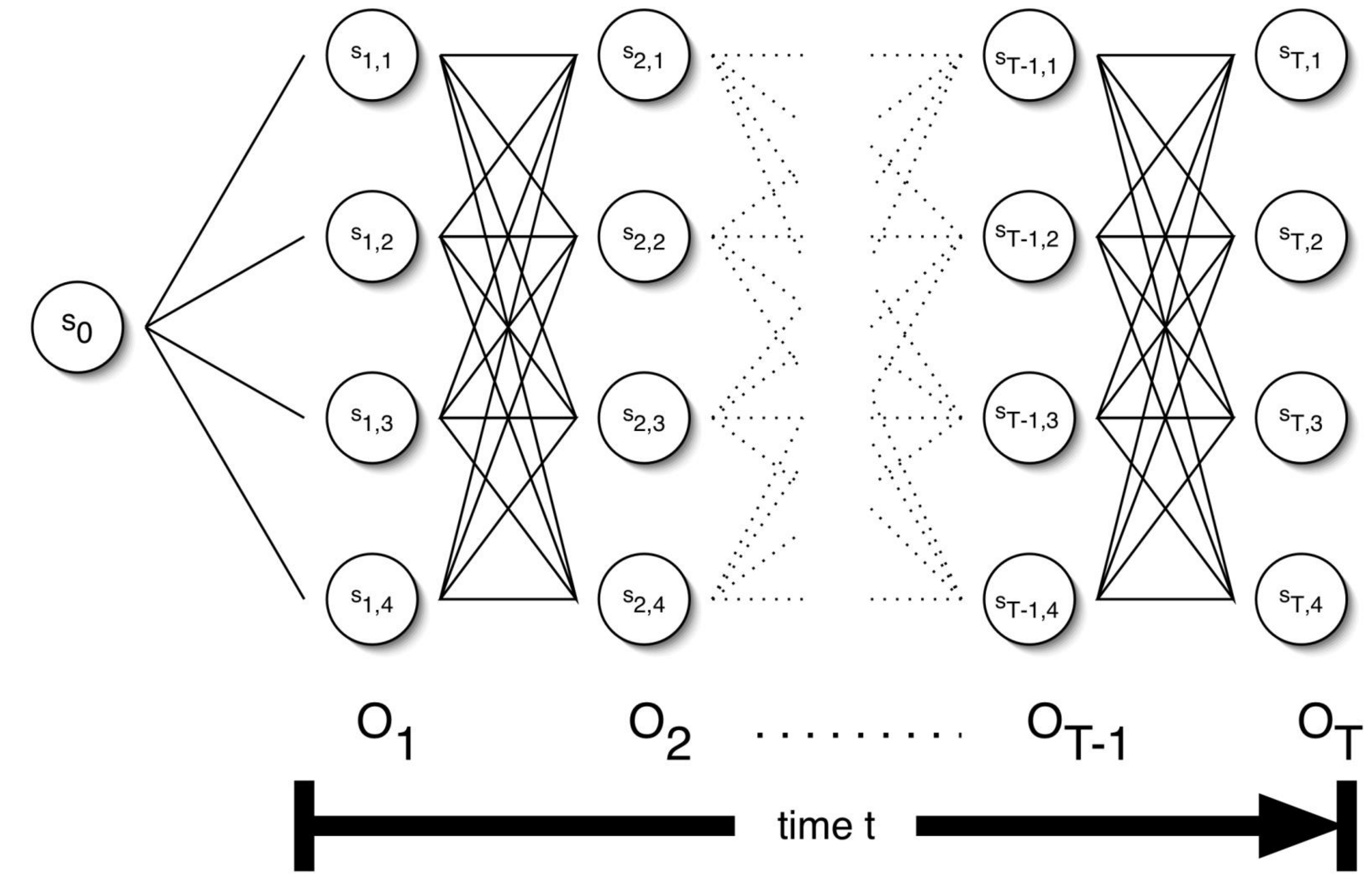
$$w_1, w_2, \ldots, w_n$$

- Np., obserwujemy sygnał akustyczny a ukryte są wypowiadane słowa

Założenie Markowa

- Prawdopodobieństwo zdarzenia w$_i$ w czasie t zależy tylko od zdarzenia w$_{i\text{-}1}$ w czasie t-1
  - Ogólnie: $$P(w_1,\ldots,w_n)=\prod_{i=2}^{n}P(w_i\mid w_1,\ldots,w_{i-1})$$
  - Założenie Markowa: $$P(w_1,\ldots,w_n)\approx\prod_{i=2}^{n}P(w_i\mid w_{i-1})$$

The Trellis (krata)



Parametry HMM

- Stany: $S=s_1,\ldots,s_n$
- Prawdopodobieństwa przejść:

  $A= a_{1,1},a_{1,2},\ldots,a_{n,n}$.

  $a_{i,j}$ – prawdop. przejścia z s_i do s_j.
- Prawdop. emisji B: zbiór funkcji $b_i(o_t)$, określających prawdop. (lub jego gęstość) obserwacji o_t wyemitowanej w stanie s_i
- Prawdopop. początkowe: π_i jest prawdopodobieństwem, że s_i jest stanem początkowym

Trzy podstawowe problemy HMM

- Problem 1 (Ewaluacja): Mając daną sekwencję obserwacji $O=o_1,\ldots,o_T$ i model HMM $\lambda=(A,B,\pi)$, obliczyć prawdopodobieństwo otrzymania tej obserwacji.
- Problem 2 (Dekodowanie): Mając daną sekwencję obserwacji $O=o_1,\ldots,o_T$ i HMM $\lambda=(A,B,\pi)$, odnaleźć sekwencję stanów (najbardziej prawdopodobną ścieżkę)

Trzy podstawowe problemy HMM

- Problem 3 (Uczenie): Obliczyć HMM

$$\lambda = (A, B, \pi)$$

na maksimum $P(O \mid \lambda)$

Problem 1: Ewaluacja

- Jak obliczyć $P(O \mid \lambda)$?
- Trzeba posumować prawdopodobieństwa dla wszystkich możliwych sekwencji stanów
- Mając T obserwacji i N stanów, mamy N^T możliwych sekwencji stanów.
- Trzeba zastosować programowanie dynamiczne.

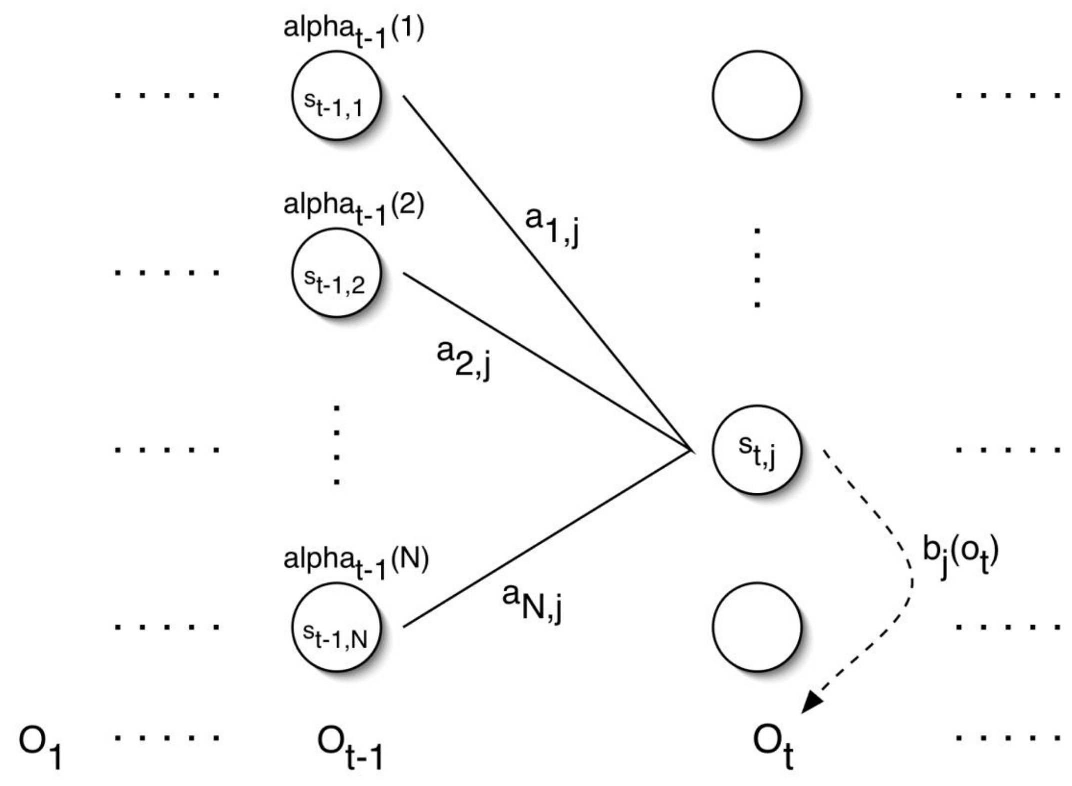
Forward Probabilities

- Mając HMM λ, czas t i stan j obliczyć prawdopodobieństwo wygenerowania częściowej sekwencji obserwacji $o_1 \ldots o_t$

$$\alpha_t(j) = P(o_1 \ldots o_t, q_t = s_j \mid \lambda)$$

Forward Probabilities

$$\alpha_t(j) = P(o_1 \ldots o_t, q_t = s_j \mid \lambda)$$



$$\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i)\, a_{ij} \right] b_j(o_t)$$

Forward Algorithm

- Inicjalizacja: $\alpha_1(i) = \pi_i b_i(o_1) \quad 1 \le i \le N$
- Indukcja:

$$\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i)\, a_{ij} \right] b_j(o_t) \quad 2 \le t \le T, 1 \le j \le N$$

- Na końcu: $P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$

  (N^2T operacji)

Backward Probabilities

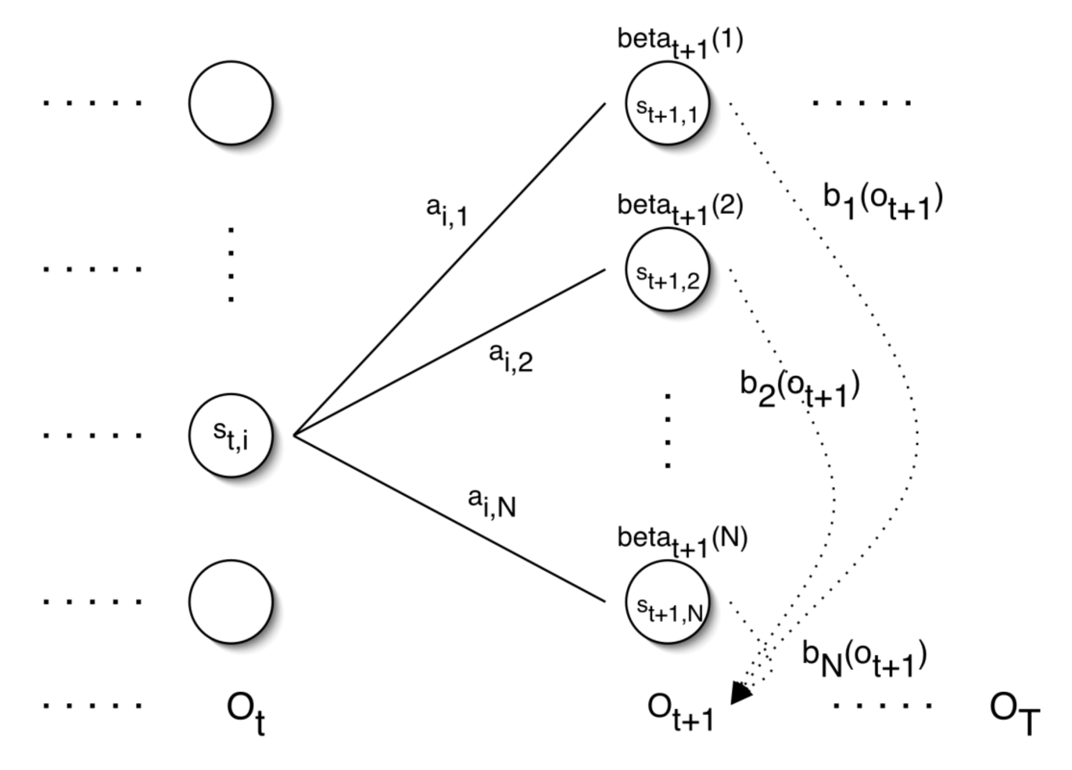
- Mając HMM λ , czas t i stan i obliczyć prawdopodobieństwo wygenerowania częściowej sekwencji obserwacji

$$o_{t+1} \ldots o_T$$

$$\beta_t(i) = P(o_{t+1} \ldots o_T \mid q_t = s_i, \lambda)$$

Backward Probabilities

$$\beta_t(i) = P(o_{t+1} \ldots o_T \mid q_t = s_i, \lambda)$$



$$\beta_t(i) = \left[\sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \right]$$

Backward Algorithm

- Inicjalizacja: $\beta_T(i) = 1, \quad 1 \le i \le N$

- Indukcja:

$$\beta_t(i) = \left[\sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \right] \; t = T-1 \ldots 1, 1 \le i \le N$$

- Na końcu:

$$P(O \mid \lambda) = \sum_{i=1}^{N} \pi_i \beta_1(i)$$

Problem 2: Dekodowanie

- Szukamy sekwencji stanów $Q=q_1 \ldots q_T$, dla których

$$Q = \arg\max_{Q'} P(Q' \mid O, \lambda)$$

Viterbi Algorithm

- Inicjalizacja: $\delta_1(i) = \pi_i b_i(o_1) \quad 1 \le i \le N$
- Indukcja:

$$\delta_t(j) = \left[\max_{1 \le i \le N} \delta_{t-1}(i)\, a_{ij}\right] b_j(o_t)$$

$$\psi_t(j) = \left\lfloor \arg\max_{1 \le i \le N} \delta_{t-1}(i)\, a_{ij} \right\rfloor \quad 2 \le t \le T, 1 \le j \le N$$

- Na końcu: $p^* = \max_{1 \le i \le N} \delta_T(i) \qquad q_T^* = \arg\max_{1 \le i \le N} \delta_T(i)$
- Ścieżka: $q_t^* = \psi_{t+1}(q_{t+1}^*) \quad t = T-1, \ldots, 1$

Problem 3: Uczenie

- Szukamy HMM λ', dla którego

$$\lambda' = \arg\max_{\lambda} P(O \mid \lambda)$$

gdzie $\lambda = (A, B, \pi)$

- Metoda Bauma-Welcha, EM: mając λ szukamy λ' , dla którego $P(O \mid \lambda') \geq P(O \mid \lambda)$
  (algorytm lokalnie zbieżny)

Problem 3: Uczenie (alg. EM)

- Dany HMM, wykonujemy alg. Forward-Backward (Expectation)
- Estymujemy nowe parametry modelu (Maximization)

(powolna, lokalna zbieżność)