Basic Problem 1 for HMM

$$1 \rightarrow 2 \rightarrow \cdots\cdots \rightarrow N \qquad \lambda = (A, B, \pi)$$

$\overline{O} = o_1 o_2 o_3 \ldots\ldots o_t \ldots\ldots o_T$ observation sequence

$\overline{q} = q_1 q_2 q_3 \ldots\ldots q_t \ldots\ldots q_T$ state sequence

- **Problem 1:** Given λ and $\overline{O}$,

  find $P(\overline{O}|\lambda)$=Prob[observing $\overline{O}$ given λ]

- **Direct Evaluation: considering all possible state sequence $\overline{q}$**

$$P(\overline{O} \mid \overline{q}, \lambda)$$

$$\Uparrow$$

$$P(\overline{O}|\lambda) = \sum_{\text{all } \overline{q}} \Big([b_{q_1}(o_1) \bullet b_{q_2}(o_2) \bullet \ldots\ldots b_{q_T}(o_T)] \bullet [\pi_{q_1} \bullet a_{q_1 q_2} \bullet a_{q_2 q_3} \bullet \ldots\ldots a_{q_{T-1} q_T}] \Big)$$

$$\Downarrow$$

$$P(\overline{q}|\lambda)$$

total number of different $\overline{q}$: N^T

huge computation requirements

Basic Problem 1 for HMM

- **Forward Procedure: defining a forward variable $\alpha_t(i)$**

$$\alpha_t(i) = P(o_1 o_2 \ldots\ldots o_t\,,\ q_t = i|\lambda)$$

$$=\text{Prob[observing } o_1 o_2 \ldots o_t\text{ , state i at time t}|\lambda]$$

- Initialization

$$\alpha_1(i) = \pi_i b_i(o_1)\ , \qquad 1 \le i \le N$$

- Induction

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

$$1 \le t \le T-1$$

$$1 \le j \le N$$

- Termination

$$P(\overline{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

See Fig. 6.5 of Rabiner and Juang

- All state sequences, regardless of how long previously, merge to the N state at each time instant t

Basic Problem 2 for HMM

- **Problem 2:** Given λ and $\overline{O} = o_1 o_2 \ldots o_T$, find a best state sequence $\overline{q} = q_1 q_2 \ldots q_T$

- **Backward Procedure :** defining a backward variable $\beta_t(i)$

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \ldots, o_T | q_t = i, \lambda)$$

$$= \text{Prob}[\text{observing } o_{t+1}, o_{t+2}, \ldots, o_T | \text{state } i \text{ at time } t, \lambda]$$

- Initialization

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

- Induction

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

$$t = T-1, T-2, \ldots, 2, 1, \quad 1 \le i \le N$$

See Fig. 6.6 of Rabiner and Juang

- **Combining Forward/Backward Variables**

$$P(\overline{O}, q_t = i | \lambda)$$

$$= \text{Prob}[\text{observing } o_1, o_2, \ldots, o_t, \ldots, o_T, q_t = i | \lambda]$$

$$= \alpha_t(i)\beta_t(i)$$

$$P(\overline{O}|\lambda) = \sum_{i=1}^{N} P(\overline{O}, q_t = i | \lambda) = \sum_{i=1}^{N} [\alpha_t(i)\beta_t(i)]$$

Basic Problem 2 for HMM

- **Approach 1 — Choosing state q_t* individually as the most likely state at time t**

  - Define a new variable $\gamma_t(i) = P(q_t = i \mid \bar{O}, \lambda)$

    $$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} = \frac{P(\bar{O}, q_t = i|\lambda)}{P(\bar{O}|\lambda)}$$

  - Solution

    $$q_t^* = \arg\max_{1 \le i \le N} [\gamma_t(i)],\ 1 \le t \le T$$

  - Problem

    maximizing the probability at each time t individually

    $\bar{q}^* = q_1^* q_2^* \ldots q_T^*$ may not be a valid sequence (e.g. $a_{q_t^* q_{t+1}^*} = 0$)

Basic Problem 2 for HMM

- **Approach 2 —Viterbi Algorithm - finding the single best sequence $\bar{q}^* = q_1^* q_2^* \ldots q_T^*$**
  - Define a new variable $\delta_t(i)$

    $$\delta_t(i) = \max_{q_1, q_2, \ldots q_{t-1}} P[q_1, q_2, \ldots q_{t-1}, q_t = i, o_1, o_2, \ldots, o_t \mid \lambda]$$

    = the highest probability along a certain single path ending at state i at time t for the first t observations, given λ
  - Induction

    $$\delta_{t+1}(j) = \max_i [\delta_t(i) a_{ij}] \bullet b_j(o_{t+1})$$
  - Backtracking

    $$\psi_t(j) = \arg\max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}]$$

    the best previous state at t–1 given at state j at time t

    keeping track of the best previous state for each j and t

Basic Problem 2 for HMM

- **Complete Procedure for Viterbi Algorithm**

  - Initialization

    $$\delta_1(i) = \pi_i b_i(o_1), \quad 1 \le i \le N$$

  - Recursion

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

    $$2 \le t \le T, \quad 1 \le j \le N$$

    $$\psi_t(j) = \arg\max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}]$$

    $$2 \le t \le T, \quad 1 \le j \le N$$

  - Termination

    $$P^* = \max_{1 \le i \le N} [\delta_T(i)]$$

    $$q_T^* = \arg\max_{1 \le i \le N} [\delta_T(i)]$$

  - Path backtracking

    $$q_t^* = \psi_{t+1}(q^*_{t+1}), \quad t = T-1, t-2, \ldots..2, 1$$

Basic Problem 3 for HMM

- **Problem 3:** Give $\overline{O}$ and an initial model $\lambda=(A,B,\pi)$, adjust λ to maximize $P(\overline{O}|\lambda)$
  - Define a new variable

$$\varepsilon_t(i, j) = P(q_t = i, q_{t+1} = j \mid \overline{O}, \lambda)$$

$$= \frac{\alpha_t(i)\, a_{ij}\, b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}[\alpha_t(i)a_{ij}\, b_j(o_{t+1})\beta_{t+1}(j)]}$$

$$= \frac{\mathrm{Prob}[\overline{O}, q_t = i, q_{t+1} = j|\lambda]}{P(\overline{O}|\lambda)}$$

See Fig. 6.7 of Rabiner and Juang

  - Recall $\gamma_t(i) = P(q_t = i \mid \overline{O}, \lambda)$

$\sum_{t=1}^{T-1}\gamma_t(i)$ = expected number of times that state i is visited in $\overline{O}$ from t = 1 to t = T–1

= expected number of transitions from state i in $\overline{O}$

$\sum_{t=1}^{T-1}\varepsilon_t(i, j)$ = expected number of transitions from state i to state j in $\overline{O}$

Basic Problem 3 for HMM

- Results

$$\bar{\pi}_i = \gamma_1(i)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \varepsilon_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\bar{b}_j(k) = \mathrm{Prob}[o_t = v_k \mid q_t = j\,] = \frac{\sum_{\substack{t=1 \\ o_t = v_k}}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

(for discrete HMM)

· **Continuous Density HMM**

$$b_j(o) = \sum_{k=1}^{M} c_{jk} N(o; \mu_{jk}, U_{jk})$$

N( ): Multi-variate Gaussian

μ_{jk}: mean vector for the k-th mixture component

U_{jk}: covariance matrix for the k-th mixture component

$\sum_{k=1}^{M} c_{jk} = 1$ for normalization

Basic Problem 3 for HMM

- **Continuous Density HMM**

  - Define a new variable

    $\gamma_{\mathbf{t}}(j, k) = \gamma_{\mathbf{t}}(j)$ but including the probability of $o_{\mathbf{t}}$ evaluated in the k-th mixture component out of all the mixture components

$$= \text{Prob}\,(q_t = j,\ m = k \mid \bar{O}, \lambda)$$

$$= \text{Prob}\,(q_t = j \mid \bar{O}, \lambda) \cdot \text{Prob}\,(m = k \mid q_t = j, \bar{O}, \lambda)$$

$$= \gamma_t(j) \cdot \text{Prob}\,(m = k \mid q_t = j, \bar{O}, \lambda)$$

$$= \left(\frac{\alpha_{\mathbf{t}}(j)\beta_{\mathbf{t}}(j)}{\sum_{j=1}^{N} \alpha_{\mathbf{t}}(j)\beta_{\mathbf{t}}(j)} \right) \left(\frac{c_{jk} N(o_{\mathbf{t}};\ \mu_{jk},\ U_{jk})}{\sum_{m=1}^{M} c_{jm} N(o_{\mathbf{t}};\ \mu_{jm},\ U_{jm})]} \right)$$

  - Results

$$\bar{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_{\mathbf{t}}(j, k)}{\sum_{t=1}^{T} \sum_{k=1}^{M} \gamma_{\mathbf{t}}(j, k)}$$

See Fig. 6.9 of Rabiner and Juang

Basic Problem 3 for HMM

- **Continuous Density HMM**

$$\bar{\mu}_{jk} = \frac{\sum_{t=1}^{T} [\gamma_t(j,k) \bullet o_t]}{\sum_{t=1}^{T} \gamma_t(j,k)}$$

$$\bar{U}_{jk} = \frac{\sum_{t=1}^{T} [\gamma_t(j,k)(o_t-\mu_{jk})(o_t-\mu_{jk})']}{\sum_{t=1}^{T} \gamma_t(j,k)}$$

- **Iterative Procedure**

$$\lambda = (A, B, \pi) \longrightarrow \bar{\lambda} = (\bar{A}, \bar{B}, \bar{\pi})$$

$$\bar{O} = o_1 o_2 \ldots o_T$$

  - It can be shown

    $P(\bar{O}|\bar{\lambda}) \geq P(\bar{O}|\lambda)$after each iteration