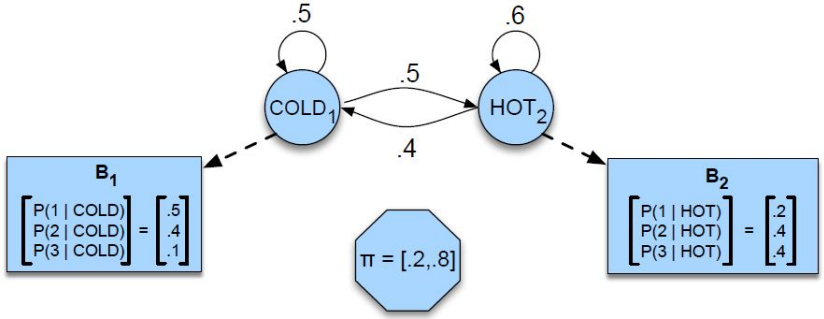
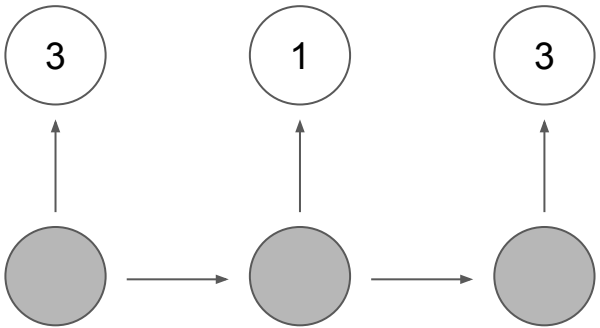
Problem 2: Decoding

Problem 2: Decoding



➔ Find the hidden state sequence that was most likely to have generated the input observation sequence.



$p(3, 1, 3)?$ **Earlier goal**

What should the goal be now?

$$\arg\max_{s_1,s_2,s_3} p(s_1, s_2, s_3 \mid o_1 = 3, o_2 = 1, o_3 = 3)?$$

Decoding: Equivalent Objective

$$\arg\max_{s_1,s_2,s_3} p(s_1, s_2, s_3 \mid o_1 = 3, o_2 = 1, o_3 = 3)?$$

$$p(s_1, s_2, s_3 \mid o_1, o_2, o_3) = \frac{p(o_1, o_2, o_3, s_1, s_2, s_3)}{p(o_1, o_2, o_3)}$$

$$\arg\max_{s_1,s_2,s_3} p(s_1, s_2, s_3 \mid o_1, o_2, o_3) = \arg\max_{s_1,s_2,s_3} p(o_1, o_2, o_3, s_1, s_2, s_3)$$

Where have we seen this quantity before?

$$p(o_{1:3}, s_{1:3})$$

$$p(o_{1:3}) = \sum\nolimits_{\forall\{s_{1:3}\}} p(o_{1:3}, s_{1:3})$$

Can you think of a brute force algorithm for decoding?

Viterbi Algorithm: Building Blocks

$$\arg\max_{s_1,s_2,s_3} p(o_1, o_2, o_3, s_1, s_2, s_3)$$

$$\text{Define: } v_t(j) = \max_{s_1,s_2,\ldots,s_{t-1}} p(s_1, s_2, \ldots, s_{t-1}, o_1, o_2, \ldots, o_t, s_t = j)$$

$$v_t(j) = \max_{s_{1:t-1}} p(s_{1:t-1}, o_{1:t}, s_t = j)$$

How do we use this to get what we want?

$$\text{Goal: } \max_{s_1,s_2,s_3} p(s_1, s_2, s_3, o_1, o_2, o_3)$$

$$\max_{s_1,s_2} p(s_1, s_2, o_1, o_2, o_3, s_3 = C) \qquad \max_{s_1,s_2} p(s_1, s_2, o_1, o_2, o_3, s_3 = H)$$

$$\max_{i=1}^{N} \max_{s_1,s_2} p(s_1, s_2, o_1, o_2, o_3, s_3 = i)$$

$$\text{Goal obtained by: } \max_{i=1}^{N} v_T(i)$$

Viterbi Algo: Recursion

$$v_t(j) = \max_{s_{1:t-1}} p(s_{1:t-1}, o_{1:t}, s_t = j)$$

$$p(s_{1:t-1}, o_{1:t}, s_t = j) = p(s_{1:t-2}, o_{1:t-1}, s_{t-1}, o_t, s_t = j)$$

$$= p(s_{1:t-2}, o_{1:t-1}, s_{t-1}) \cdot p(o_t, s_t = j \mid s_{1:t-2}, o_{1:t-1}, s_{t-1})$$

$$= p(s_{1:t-2}, o_{1:t-1}, s_{t-1}) \cdot p(s_t = j \mid s_{1:t-2}, o_{1:t-1}, s_{t-1}) \cdot p(o_t \mid s_t = j, s_{1:t-2}, o_{1:t-1}, s_{t-1})$$

$$= p(s_{1:t-2}, o_{1:t-1}, s_{t-1}) \cdot p(s_t = j \mid s_{t-1}) \cdot p(o_t \mid s_t = j)$$

$$p(s_{1:t-1}, o_{1:t}, s_t = j) = p(s_{1:t-2}, o_{1:t-1}, s_{t-1}) \cdot p(s_t = j \mid s_{t-1}) \cdot p(o_t \mid s_t = j)$$

Viterbi Algo: Recursion

$$p(s_{1:t-1}, o_{1:t}, s_t = j) = p(s_{1:t-2}, o_{1:t-1}, s_{t-1}) \cdot p(s_t = j \mid s_{t-1}) \cdot p(o_t \mid s_t = j)$$

$$\max_{s_{1:t-1}} p(s_{1:t-1}, o_{1:t}, s_t = j)$$

$$= \max_{i=1}^{N} \max_{s_{1:t-2}} p(s_{1:t-2}, o_{1:t-1}, s_{t-1} = i) \cdot p(s_t = j \mid s_{t-1} = i) \cdot p(o_t \mid s_t = j)$$

$$v_t(j) = \max_{i=1}^{N} v_{t-1}(i) \cdot t_{ij} \cdot b_j(o_t)$$

Viterbi Algorithm

1. **Initialization:**

$$\begin{aligned} v_1(j) &= \pi_j b_j(o_1) \qquad 1 \le j \le N \\ bt_1(j) &= 0 \qquad\qquad\quad 1 \le j \le N \end{aligned}$$

2. **Recursion**

$$\begin{aligned} v_t(j) &= \max_{i=1}^{N} v_{t-1}(i)\, a_{ij}\, b_j(o_t); \quad 1 \le j \le N, 1 < t \le T \\ bt_t(j) &= \operatorname*{argmax}_{i=1}^{N} v_{t-1}(i)\, a_{ij}\, b_j(o_t); \quad 1 \le j \le N, 1 < t \le T \end{aligned}$$

3. **Termination:**

$$\begin{aligned} \text{The best score:} \quad P* &= \max_{i=1}^{N} v_T(i) \\ \text{The start of backtrace:} \quad q_T* &= \operatorname*{argmax}_{i=1}^{N} v_T(i) \end{aligned}$$

Source: SLP, Dan Jurafsky

Viterbi Algorithm: Pseudocode

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do**
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$

$$v_1(j) = \pi_j b_j(o_1) \qquad 1 \le j \le N$$
$$bt_1(j) = 0 \qquad 1 \le j \le N$$

for each time step t **from** 2 **to** T **do** ; recursion step
  for each state s **from** 1 **to** N **do**
    $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
    $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$

$$v_t(j) = \max_{i=1}^{N} v_{t-1}(i)\, a_{ij}\, b_j(o_t);$$
$$bt_t(j) = \operatorname{argmax}_{i=1}^{N} v_{t-1}(i)\, a_{ij}\, b_j(o_t)$$

$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step

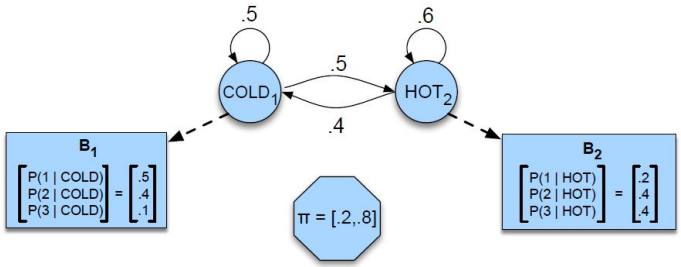
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step

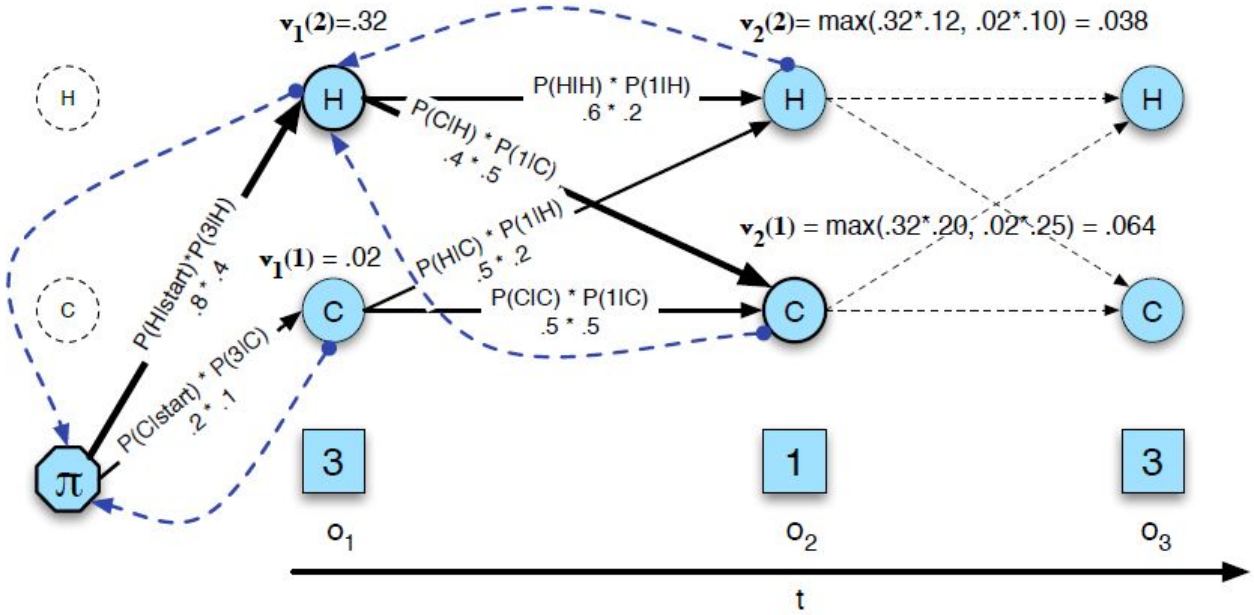
$$P* = \max_{i=1}^{N} v_T(i)$$
$$q_T* = \operatorname{argmax}_{i=1}^{N} v_T(i)$$

$bestpath \leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time

Source: SLP, Dan Jurafsky

Viterbi Trellis



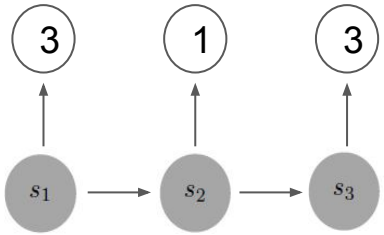


Source: SLP, Dan Jurafsky

Review

Likelihood

Decoding



$$p(3, 1, 3)?$$

$$\arg\max_{s_1, s_2, s_3} p(s_1, s_2, s_3 \mid o_1 = 3, o_2 = 1, o_3 = 3)?$$

$$\arg\max_{s_1, s_2, s_3} p(o_1, o_2, o_3, s_1, s_2, s_3)$$

$$\alpha_t(j) = p(o_1, o_2, \ldots, o_t, s_t = j)$$

$$v_t(j) = \max_{s_{1:t-1}} p(s_{1:t-1}, o_{1:t}, s_t = j)$$

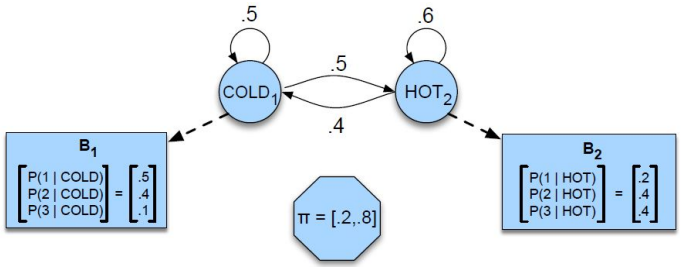
$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) \cdot p(s_t = j \mid s_{t-1} = i) \cdot p(o_t \mid s_t = j)$$

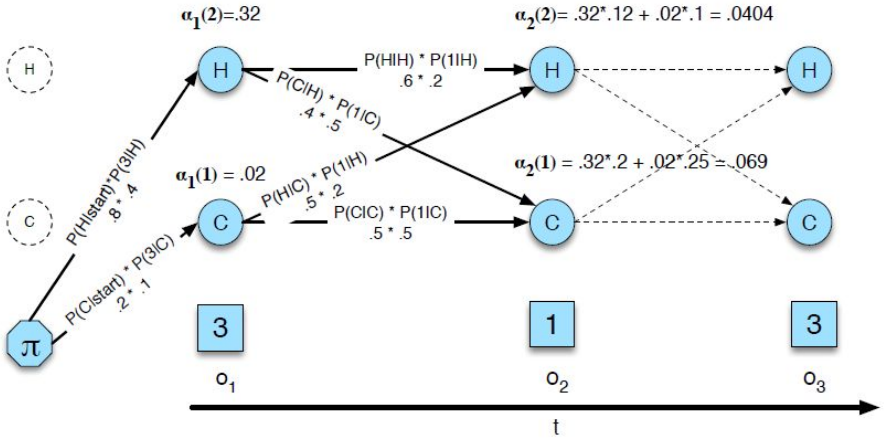
$$v_t(j) = \max_{i=1}^{N} v_{t-1}(i) \cdot t_{ij} \cdot b_j(o_t)$$

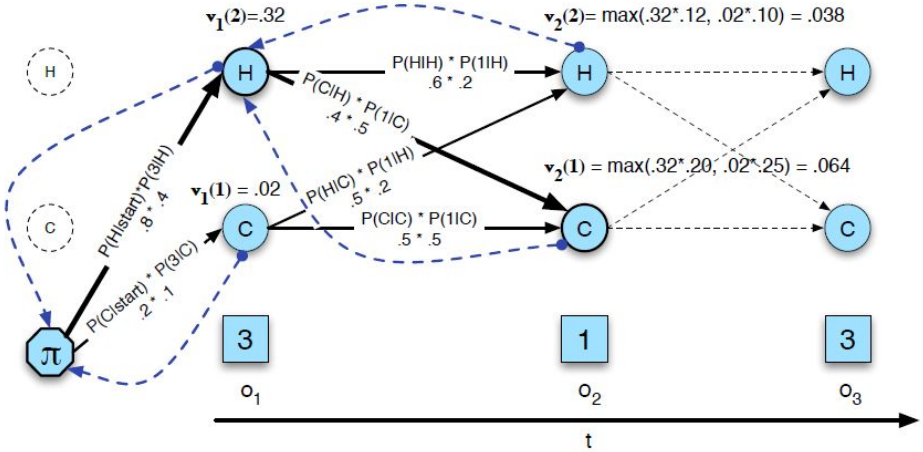
$$p(O) = \sum_{i=1}^{N} \alpha_T(i)$$

$$\max_{i=1}^{N} v_T(i)$$

Trellis: Forward vs. Viterbi







DP = recursion + memoization?

Source: SLP, Dan Jurafsky