Viterbi's Algorithm

Nirdosh Bhatnagar

- For a given observation sequence o, the corresponding state sequence q, has to be determined which is optimal in some predetermined sense.
- That is, this problem requires the determination of the hidden-state sequence that will most likely produce the specified observation sequence.
- Let the length of each of the hidden-state and output sequences be T. Note that

$$o = (o_1, o_2, \ldots, o_T), \quad \text{and} \quad q = (q_1, q_2, \ldots, q_T)$$

- Also let Q be the random hidden-state sequence vector and O be the corresponding random output sequence vector. Each of these random vectors is of length T.
- A popular criterion is to find a hidden-state sequence (path) which maximizes the probability $P(Q = q \mid O = o)$. This is equivalent to maximizing $P(Q = q, O = o)$.
- A technique to finding such sequence uses the methods of dynamic programming. The actual technique is called the Viterbi algorithm, named after its inventor, Andrew J. Viterbi.

Outline of the Algorithm

- Let S be the set of states, where $|S| = N$. That is, the total number of possible states is equal to N.
- Define $\delta_t(i)$ as the probability of the highest probability sequence (path) at time $t \in \mathbb{P}$, which accounts for the first t output symbols, and which ends in state $s_i \in S$. Thus

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1}} P(q_1, q_2, \ldots, q_{t-1}, q_t = s_i; o_1, o_2, \ldots, o_t)$$

- Use of induction yields

$$\delta_{t+1}(j) = \left\{ \max_i \delta_t(i) a_{ij} \right\} b_j(o_{t+1})$$

- The state sequence can be kept track of, via an array $\psi_t(j)$. That is, $\psi_t(j)$ keeps track of the state that maximizes $\delta_{t-1}(i) a_{ij}$ over i, which is the optimal previous state.
- Thus $\delta_{t+1}(i)$ can be computed recursively, and the optimal path can be determined via backtracking from time T. The Viterbi algorithm is as follows.

Step 1: Initialization:

$$\begin{aligned}\delta_1(i) &= \pi_i b_i(o_1), \quad 1 \le i \le N \\ \psi_1(i) &= 0, \quad 1 \le i \le N\end{aligned}$$

Step 2: Recursion:

$$\begin{aligned}\delta_t(i) &= \max_{1 \le k \le N} \{\delta_{t-1}(k)\, a_{ki}\}\, b_i(o_t), \quad 2 \le t \le T, \quad 1 \le i \le N \\ \psi_t(i) &= \arg\max_{1 \le k \le N} \{\delta_{t-1}(k)\, a_{ki}\}, \quad 2 \le t \le T, \quad 1 \le i \le N\end{aligned}$$

Step 3: Termination:

$$\begin{aligned}P^* &= \max_{1 \le k \le N} \delta_T(k) \\ q_T^* &= \arg\max_{1 \le k \le N} \delta_T(k)\end{aligned}$$

Step 4: Optimal state sequence (path) backtracking:

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

□

Observe that, the above algorithm involves multiplication of probabilities (which are numbers between zero and unity). This might result in underflow in the computations. The above Viterbi algorithm can be modified by taking logarithms. This eliminates the need for multiplications.

Alternative Viterbi's Algorithm

Step 0: Preprocessing:

$$\begin{aligned}\widehat{\pi}_i &= \log(\pi_i), \quad 1 \le i \le N \\ \widehat{b}_i(o_t) &= \log\{b_i(o_t)\}, \quad 1 \le i \le N, 1 \le t \le T \\ \widehat{a}_{ij} &= \log(a_{ij}), \quad 1 \le i, j \le N\end{aligned}$$

Step 1: Initialization:

$$\begin{aligned}\widehat{\delta}_1(i) &= \log\{\delta_1(i)\} = \widehat{\pi}_i + \widehat{b}_i(o_1), \quad 1 \le i \le N \\ \psi_1(i) &= 0, \quad 1 \le i \le N\end{aligned}$$

Step 2: Recursion:

$$\begin{aligned}\widehat{\delta}_t(i) &= \log(\delta_t(i)) = \max_{1 \le k \le N} \left\{\widehat{\delta}_{t-1}(k) + \widehat{a}_{ki}\right\} + \widehat{b}_i(o_t), \quad 2 \le t \le T, \quad 1 \le i \le N \\ \psi_t(i) &= \arg\max_{1 \le k \le N} \left\{\widehat{\delta}_{t-1}(k) + \widehat{a}_{ki}\right\}, \quad 2 \le t \le T, \quad 1 \le i \le N\end{aligned}$$

Step 3: Termination:

$$\begin{aligned} P^* &= \max_{1 \le k \le N} \widehat{\delta}_T(k) \\ q_T^* &= \arg\max_{1 \le k \le N} \widehat{\delta}_T(k) \end{aligned}$$

Step 4: Optimal state sequence (path) backtracking:

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

□

Example

$S = \{1, 2, 3\}$ = set of hidden states
$V = \{G, R\}$ = set of output states
$A = [a_{ij}]$ = hidden-state transition matrix (size $|S| \times |S| = 3 \times 3$)
B = emission matrix (size $|S| \times |V| = 3 \times 2$)
π = initial distribution (size $1 \times |S| = 1 \times 3$)

$$A = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0.5 & 0.2 & 0.3 \\ 0.4 & 0.1 & 0.5 \end{bmatrix} \end{array}$$

$$B = \begin{array}{c c} & \begin{array}{cc} G & R \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \\ 1/4 & 3/4 \end{bmatrix} \end{array}$$

$$\pi = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ 1 & \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix} \end{array}$$

$$\begin{aligned} b_1(G) &= 1/2, \qquad b_1(R) = 1/2 \\ b_2(G) &= 2/3, \qquad b_2(R) = 1/3 \\ b_3(G) &= 1/4, \qquad b_3(R) = 3/4 \end{aligned}$$

Observable sequence $= O = (o_1, o_2, o_3) = (R, G, R)$
Hidden state sequence $= q = (q_1, q_2, q_3) = ?$
That is, determine the hidden state sequence with maximum probability.

Solution

Initialization:

$t = 1$, $o_1 = R$

$$\delta_1(i) = \pi_i b_i(o_1) = \pi_i b_i(R), \quad 1 \le i \le 3$$

$$\begin{aligned} \delta_1(1) &= \pi_1 b_1(R) = (1/3)(1/2) = 1/6 \\ \delta_1(2) &= \pi_2 b_2(R) = (1/3)(1/3) = 1/9 \\ \delta_1(3) &= \pi_3 b_3(R) = (1/3)(3/4) = 1/4 \end{aligned}$$

$$\psi_1(i) = 0, \quad 1 \le i \le 3$$
$$\psi_1(1) = \psi_1(2) = \psi_1(3) = 0$$

Recursion:

$t = 2$, $o_2 = G$

$$\delta_2(i) = \max_{1 \le k \le 3} \{\delta_1(k) a_{ki}\} b_i(o_2), \quad 1 \le i \le 3$$

$$\begin{aligned}
\delta_2(1) &= \max\{\delta_1(1) a_{11},\ \delta_1(2) a_{21},\ \delta_1(3) a_{31}\} b_1(G) \\
&= \max\{(1/6)(0.3),\ (1/9)(0.5),\ (1/4)(0.4)\} (1/2) \\
&= (1/4)(0.4)(1/2) = 0.05 \\
\psi_2(1) &= 3
\end{aligned}$$

$$\begin{aligned}
\delta_2(2) &= \max\{\delta_1(1) a_{12},\ \delta_1(2) a_{22},\ \delta_1(3) a_{32}\} b_2(G) \\
&= \max\{(1/6)(0.6),\ (1/9)(0.2),\ (1/4)(0.1)\} (2/3) \\
&= (1/6)(0.6)(2/3) = 1/15 \\
\psi_2(2) &= 1
\end{aligned}$$

$$\begin{aligned}
\delta_2(3) &= \max\{\delta_1(1) a_{13},\ \delta_1(2) a_{23},\ \delta_1(3) a_{33}\} b_3(G) \\
&= \max\{(1/6)(0.1),\ (1/9)(0.3),\ (1/4)(0.5)\} (1/4) \\
&= (1/4)(0.5)(1/4) = 0.03125 \\
\psi_2(3) &= 3
\end{aligned}$$

$t = 3$, $o_3 = R$

$$\delta_3(i) = \max_{1 \le k \le 3} \{\delta_2(k) a_{ki}\} b_i(o_3), \quad 1 \le i \le 3$$

$$\begin{aligned}
\delta_3(1) &= \max\{\delta_2(1) a_{11},\ \delta_2(2) a_{21},\ \delta_2(3) a_{31}\} b_1(R) \\
&= \max\{(0.05)(0.3),\ (1/15)(0.5),\ (1/32)(0.4)\} (1/2) \\
&= (1/15)(0.5)(1/2) = 1/60 \\
\psi_3(1) &= 2
\end{aligned}$$

$$\begin{aligned}
\delta_3(2) &= \max\{\delta_2(1) a_{12},\ \delta_2(2) a_{22},\ \delta_2(3) a_{32}\} b_2(R) \\
&= \max\{(0.05)(0.6),\ (1/15)(0.2),\ (1/32)(0.1)\} (1/3) \\
&= (0.05)(0.6)(1/3) = 0.01 \\
\psi_3(2) &= 1
\end{aligned}$$

$$\begin{aligned}
\delta_3(3) &= \max\{\delta_2(1) a_{13},\ \delta_2(2) a_{23},\ \delta_2(3) a_{33}\} b_3(R) \\
&= \max\{(0.05)(0.1),\ (1/15)(0.3),\ (1/32)(0.5)\} (3/4) \\
&= (1/15)(0.3)(3/4) = 0.015 \\
\psi_3(3) &= 2
\end{aligned}$$

Termination:

$$P^* = \max_{1 \leq k \leq 3} \delta_3(k) = \max\{1/60, 0.01, 0.015\} = 1/60$$

$$q_3^* = \arg\max_{1 \leq k \leq 3} \delta_3(k) = 1$$

Backtracking:

$$q_t^* = \psi_{t+1}\left(q_{t+1}^*\right), \qquad t = 2, 1$$

$$\begin{aligned} q_2^* &= \psi_3(q_3^*) = \psi_3(1) = 2 \\ q_1^* &= \psi_2(q_2^*) = \psi_2(2) = 1 \end{aligned}$$

In summary, for the observation $O = (o_1, o_2, o_3) = (R, G, R)$,
Maximum probability $P^* = 1/60$
The corresponding hidden state sequence is equal to $q^* = (q_1^*, q_2^*, q_3^*) = (1, 2, 1)$. $\square$

$\delta_1(i) = \pi_i b_i(R)$

$i=1$: $\frac{1}{3}\cdot\frac{1}{2} = \frac{1}{6}$

$\psi_1(1) = 0$

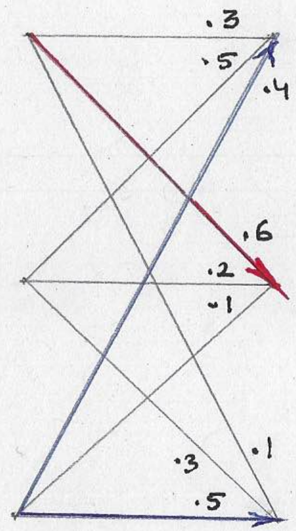
$i=2$: $\frac{1}{3}\cdot\frac{1}{3} = \frac{1}{9}$

$\psi_1(2) = 0$

$i=3$: $\frac{1}{3}\cdot\frac{3}{4} = \frac{1}{4}$

$\psi_1(3) = 0$

$t=1$



$\delta_2(i) =$

$\max\limits_{1\le k\le 3}\left[\delta_1(k)\, a_{ki}\right] b_i(\overset{G}{\cancel{B}})$

$\left[\frac{1}{4}(0.4)\right]\frac{1}{2} = 0.05$

$\psi_2(1) = 3$

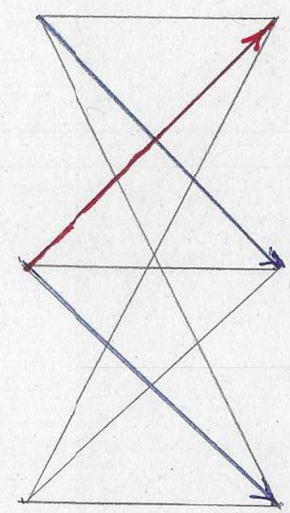
$\left[\frac{1}{6}(0.6)\right]\frac{2}{3} = \frac{1}{15}$

$\psi_2(2) = 1$

$\left[\frac{1}{4}(0.5)\right]\frac{1}{4} = \frac{1}{32}$

$\psi_2(3) = 3$

$t=2$



$\delta_3(i) =$

$\max\limits_{1\le k\le 3}\left[\delta_2(k)\, a_{ki}\right] b_i(R)$

$\left[\frac{1}{15}(0.5)\right]\frac{1}{2} = \frac{1}{60}$

$\psi_3(1) = 2$

$\left[0.05(0.6)\right]\frac{1}{3} = 0.01$

$\psi_3(2) = 1$

$\left[0.0\bar{6}(0.3)\right]\frac{3}{4} = 0.015$

$\psi_3(3) = 2$

$t=3$

VITERBI'S TRELLIS DIAGRAM