STAT 534 Lecture 8

May 16, 2013

Hidden Markov Models Summary

©Marina Meilă

mmp@stat.washington.edu

Reading: Article by Lawrence Rabiner "A tutorial on hidden markov models and selected applications in speech recognition"; [Optional CRLS Problem 15–5, page 367]

Notation (consistent with the Rabiner article)	
$q_t \in \{1, \ldots, N\}$	the state at time t
$O_t \in \{1, \ldots, M\}$	the output at time t
$A = [a_{ij}]_{i,j=1}^{N}$	the transition matrix
$B = [b_i(k)]_{i=1}^{N}{}_{k=1}^{M}$	the emission matrix
$\pi = [\pi_i]_{i=1}^{M}$	the initial state distribution

1 The Forward-Backward Algorithm

$$
\begin{aligned}
\alpha_t(i) &= P[\, O_{1:t},\ q_t = i\,] \\
\beta_t(i) &= P[\, O_{t+1:T} \,|\, q_t = i\,]
\end{aligned}
$$

The Forward pass

$$
\begin{aligned}
\alpha_1(i) &= \pi_i b_i(O_1) \\
\alpha_t(i) &= \sum_{j=1}^{N} \alpha_{t-1}(j) a_{ji} b_i(O_t)
\end{aligned}
$$

The Backward pass

$$
\beta_T(i) = 1
$$

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

The likelihood of the outputs

$$P[\,O_{1:T}\,] \;=\; \sum_{i=1}^{N} \alpha_t(i)\beta_t(i) \;=\; \sum_{i=1}^{N} \alpha_T(i)$$

The probability of a state or transition given the data

$$\begin{aligned}
\gamma_t(i) &= P[\,q_t = i \,|\, O_{1:T}\,] \;=\; \frac{\alpha_t(i)\beta_t(i)}{P[\,O_{1:T}\,]} \\
\xi_t(i,j) &= P[\,q_t = i,\; q_{t+1} = j \,|\, O_{1:T}\,] \;=\; \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P[\,O_{1:T}\,]}
\end{aligned}$$

2 The Baum-Welch Algorithm

Expectation step: Compute $\alpha,\ \beta,\ \gamma,\ \xi$ from the current parameter estimates $A,\ B,\ \pi$

Maximization step: Reestimate parameters by

$$\begin{aligned}
\pi_i &= \gamma_1(i) \\
a_{ij} &= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \\
b_i(k) &= \frac{\sum_{t:\,O_t=k} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(i)}
\end{aligned}$$

3 The Viterbi Algorithm

Notation	
$\delta_t(i)$	the optimum cost of a state sequence $[q_{1:t-1},\, q_t = i]$
	$= \max_{q_{1:t-1}} P[q_{1:t-1},\, q_t = i,\, O_{1:t}]$
$\psi_t(i)$	pointer for backtracking = q_{t-1} in the optimal state sequence

Initialization

$$\begin{aligned} \delta_1(i) &= \pi_i b_i(O_1) \\ \psi_1(i) &= \text{undefined} \end{aligned}$$

Recursion

$$\begin{aligned} \delta_t(i) &= \max_{j=1,\ldots N} [\delta_{t-1}(j) a_{ji}] b_i(O_t) \\ \psi_t(i) &= \operatorname*{argmax}_{j=1,\ldots N} [\delta_{t-1}(j) a_{ji}] \end{aligned}$$