Linguistics 696f
Hammond
Spring '03

Algorithms for Hidden Markov Models continued...

A. Overview

(1)
a. Forward probabilities
b. Backward probabilities
c. Training a simple Markov chain
d. Baum-Welch

B. Forward probabilities

(2) $\alpha_i(t)$ the probability of producing $w_{1,t-1}$ and ending up in s^i.

(3) $\alpha_i(t) \stackrel{\text{def}}{=} \text{p}(w_{1,t-1}, S_t = s^i)$, where $t > 1$

(4) $\alpha_i(1) = \pi_i$

(5) $\alpha_j(t+1) = \sum_{i=1}^{\sigma} \alpha_i(t)\text{p}(s^i \stackrel{w_t}{\rightarrow} s^j)$

C. Backward probabilities

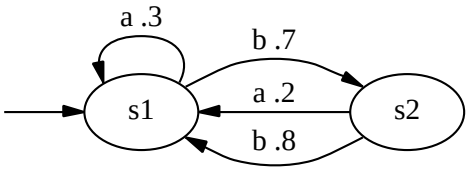
(6) $\beta_i(t)$ is the overall probability of producing $w_{t,n}$ where the HMM is in state s^i at time t.

(7) $\beta_i(n+1) = \text{p}(\epsilon | S_{n+1} = s^i) = 1$

(8) $\beta_i(t-1) = \sum_{j=1}^{\sigma} \text{p}(s^i \stackrel{w_{t-1}}{\rightarrow} s^j)\beta_j(t)$

D. Training a simple Markov chain

(9) A simple Markov chain



(10) $$\mathrm{p}_e(s^i \xrightarrow{w^k} s^j) = \frac{|s^i \xrightarrow{w^k} s^j|}{\sum_{l=1,m=1}^{\sigma,\omega} |s^i \xrightarrow{w^m} s^l|}$$

(11) $$s^1 \xrightarrow{a} s^1 \xrightarrow{b} s^2 \xrightarrow{a} s^1 \xrightarrow{b} s^2 \xrightarrow{a} s^1 \xrightarrow{b} s^2 \xrightarrow{b} s^1$$

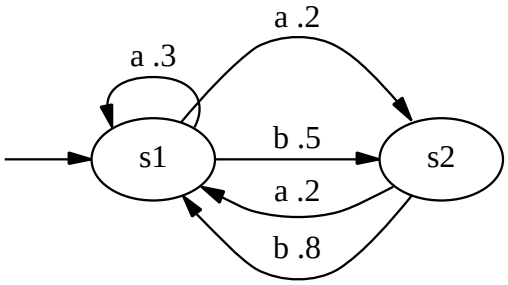
(12)

arc	count	probability
$s^1 \xrightarrow{a} s^1$	1	.25
$s^1 \xrightarrow{b} s^2$	3	.75
$s^2 \xrightarrow{a} s^1$	2	.66
$s^2 \xrightarrow{b} s^1$	1	.33

E. Baum-Welch

(13) $$|s^i \xrightarrow{w^k} s^j| = \frac{1}{\mathrm{p}(w_{1,n})} \sum_{t=1}^{n} \alpha_i(t)\mathrm{p}(s^i \xrightarrow{w^k} s^j)\beta_j(t+1)$$

(14) A simple Hidden Markov Model



(15) We will train the HMM above with *aabb*. We first calculate the values for α and β.

(16)

	$\alpha(1)$	$\alpha(2)$	$\alpha(3)$	$\alpha(4)$	$\alpha(5)$
s^1	1	.3	.13	.048	.052
s^2	0	.2	.06	.065	.024

(17)

	$\beta(1)$	$\beta(2)$	$\beta(3)$	$\beta(4)$	$\beta(5)$
s^1	.076	.2	.4	.5	1
s^2	0	.08	.4	.8	1

(18)

$$
\begin{array}{lcll}
|s^i \xrightarrow{w^k} s^j| & = & \frac{1}{\mathrm{p}(w_{1,n})} \sum_{t=1}^{n} \alpha_i(t)\mathrm{p}(s^i \xrightarrow{w^k} s^j)\beta_j(t+1) & \\
\hline
|s^1 \xrightarrow{a} s^1| & = & \frac{1}{.074}[(1 \times .3 \times .2) + (.3 \times .3 \times .4)] & = \quad 1.2972 \\
|s^1 \xrightarrow{a} s^2| & = & \frac{1}{.074}[(1 \times .2 \times .08) + (.3 \times .2 \times .4)] & = \quad .5405 \\
|s^1 \xrightarrow{b} s^2| & = & \frac{1}{.074}[(.13 \times .5 \times .8) + (.048 \times .5 \times 1)] & = \quad 1.027 \\
|s^2 \xrightarrow{a} s^1| & = & \frac{1}{.074}[(0 \times .2 \times .2) + (.2 \times .2 \times .4)] & = \quad .2162 \\
|s^2 \xrightarrow{b} s^1| & = & \frac{1}{.074}[(.06 \times .8 \times .5) + (.065 \times .8 \times 1)] & = \quad 1.027
\end{array}
$$

(19)

$\lvert s^i \xrightarrow{w^k} s^j \rvert$	p_e
$\lvert s^1 \xrightarrow{a} s^1 \rvert$	.45
$\lvert s^1 \xrightarrow{a} s^2 \rvert$	.19
$\lvert s^1 \xrightarrow{b} s^2 \rvert$	.36
$\lvert s^2 \xrightarrow{a} s^1 \rvert$	.17
$\lvert s^2 \xrightarrow{b} s^1 \rvert$	.83

References

Charniak, Eugene. 1993. *Statistical Language Learning*. Cambridge: MIT Press.

Hammond, Michael, 2003. Statistical natural language processing. U. of Arizona.

Hopcroft, J.E., & J.D. Ullman. 1979. *Introduction to Automata Theory, Languages, and Computation*. Reading: Addison-Wesley.