1

"I hate it when ___ goes ___ like that EOS"

We solve this problem using the trigram model.

Assume that the first blank is referred to as b_1 and the second is b_2.

We wish to maximise over all x and y

$Pr(b_1 = x, b_2 = y \mid \text{I hate it when } b_1 \text{ goes } b_2 \text{ like that } EOS)$.

This is maximised by

$$\frac{Pr(\text{I hate it when } x \text{ goes } y \text{ like that } EOS)}{\sum_{(m,n)} Pr(\text{I hate it when } m \text{ goes } n \text{ like that } EOS)}$$

In the term above, the since the denominator is a constant term not dependent on x and y, we can maximise it by just maximising the numerator.

$Pr(\text{I hate it when } x \text{ goes } y \text{ like that } EOS) =$
$Pr(\text{it} \mid \text{I hate}) * Pr(\text{when} \mid \text{hate it}) * Pr(x \mid \text{it when}) * Pr(\text{goes} \mid \text{when } x) *$
$Pr(y \mid x \text{ goes}) * Pr(\text{like} \mid \text{goes } y) * Pr(\text{that} \mid y \text{ like}) * Pr(\text{EOS} \mid \text{like that})$.

Since the terms not containing references to x or y are constant, we are left with the task of maximising

$Pr(x|\text{it when}) * Pr(\text{goes}|\text{when } x) * Pr(y|x \text{ goes}) * Pr(\text{like}|\text{goes } y) * Pr(\text{that}|y \text{ like})$

2a

$Pr(k_0, \ldots, k_n, k_{n+1}) =$

$\Sigma_i Pr(k_0, \ldots, k_{n+1}, s_n = i)$

"since we could end up in any state, i"

$= \Sigma_j \ \Sigma_i Pr(k_0, \ldots, k_{n+1}, s_n = i, s_{n+1} = j)$

"since the preceeding state before i could be any j"

$= \Sigma_j \ \Sigma_i Pr(k_{n+1}, s_{n+1} = j | k_0, \ldots, k_n, s_n = i) Pr(k_0, \ldots, k_n, s_n = i)$

"using conditional probability"

$= \Sigma_j \ \Sigma_i \alpha(i, n) Pr(k_{n+1}, s_{n+1} = j | k_0, \ldots, k_n, s_n = i)$

"definition of α"

$$= \Sigma_j \ \Sigma_i \left(\begin{array}{l} \alpha(i,n) Pr(s_{n+1} = j | k_0, \ldots, k_n, s_n = i) * \\ Pr(k_{n+1} | s_{n+1} = j, k_0, \ldots, k_n, s_n = i) \end{array} \right)$$

"chain rule for conditional probability"

$= \Sigma_j \ \Sigma_i \alpha(i,n) Pr(s_{n+1} = j | s_n = i) Pr(k_{n+1} | s_{n+1} = j)$
"since observation is dependent only on current state, and current state is dependent only on earlier state"

$= \Sigma_j \ \Sigma_i \alpha(i,n) p(i,j) b(j,k)$
"definition of p and b"

$= \Sigma_j \ \alpha(j, n+1)$
"definition of α"

2b

$\alpha(i,0) =$
$Pr(k_0, s_0 = i) =$
$Pr(k_0 | s_0 = i) Pr(s_0 = i) =$
$b(k_0, i)\pi(i)$

Found an error? Please contact kswadi@cs.princeton.edu