COMS E6998 Homework 2

Mengqi Zong $< mz2326@columbia.edu >$

April 12, 2012

Question 1

Initialization: Set $q^0(y|w_1)$ and $r^0(w_2|y)$ to some initial values (e.g. random initial values).

Algorithm: For $t = 1...T$:

1. For $i = 1...n$, and $y = 1...K$, calculate

$$\delta(y|i) = p(y|w_1^i, w_2^i; \underline{\theta}^{t-1}) = \frac{q^{t-1}(y|w_1^i)r(w_2^i|y)}{\sum_y q^{t-1}(y|w_1^i)r(w_2^i|y)}$$

2. Recalculate the parameters:

$$q^t(y|w_1) = \frac{\sum_{i:w_1^i=w_1} \delta(y|i)}{\sum_{i=1}^n [[w_1^i = w_1]]}$$
$$r^t(w_2|y) = \frac{\sum_{i:w_2^i=w_2} \delta(y|i)}{\sum_{i=1}^n \delta(y|i)}$$

Question 2

Question 2(a)

$$P(y_2 = 1, y_3 = 2, y_4 = 1|x; \theta) = \frac{\alpha[2,1] \times t(2,1) \times e(x_3, 2) \times t(1,2) \times e(x_4|1) \times \beta[4,1]}{\sum_s \alpha[n, s]}$$

Note that n is the length of the sequence of x.

Question 2(b)

$$P(y_2 = 1, y_5 = 1|x;\theta) = \frac{\sum_{y_3}\sum_{y_4}\alpha[2,1]t(y_3,1)e(x_3,y_3)t(y_4,y_3)e(x_4,y_4)t(1,y_4)e(x_5,1)\beta[5,1]}{\sum_s \alpha[n,s]}$$

Note that n is the length of the sequence of x.

Question 2(c)

Here is the modified definition of the forward and backward terms:

$$\begin{aligned}
\alpha'[j,s] &= \max_{s_1 \ldots s_{j-1}} [t(s_1)e(x_1|s_1)(\prod_{k=2}^{j-1} t(s_k|s_{k-1})e(x_k|s_k))t(s|s_{j-1})e(x_J|s)] \\
\beta'[j,s] &= \max_{s_{j+1} \ldots s_m} [t(s_{j+1}|s)e(x_{j+1}|s_{j+1})(\prod_{k=j+2}^{m} t(s_k|s_{k-1})e(x_k|s_k))]
\end{aligned}$$

Here is the recursive method. For $j = 2...m$:

$$\begin{aligned}
\alpha'[j,s] &= \max_{s' \in \{1..k\}} (\alpha'[j-1,s'] \times t(s|s') \times e(x_j|s)) \\
\beta'[j,s] &= \max_{s' \in \{1..k\}} (\beta'[j+1,s'] \times t(s'|s) \times e(x_{j+1}|s'))
\end{aligned}$$

The probability we want to calculate is:

$$\begin{aligned}
\max_{y:y_j=p} p(y|x;\theta) &= \alpha'[j,p] \times \beta'[j,p] \\
\max_{y:y_3=1} p(y|x;\theta) &= \alpha'[3,1] \times \beta'[3,1]
\end{aligned}$$

Question 3

Question 3(a)

$$
\begin{aligned}
\alpha[j,a] &= \sum_{a_1 \ldots a_{j-1}} [d(a_1|a_0)t(f_1|e_{a_1}))(\prod_{k=2}^{j-1} d(a_k|a_{k-1})t(f_k|e_{a_k}))d(a|a_{j-1})t(f_j|e_{a_j})] \\
\beta[j,a] &= \sum_{a_{j+1} \ldots a_m} [d(a_{j+1}|a)e(f_{j+1}|e_{a_{j+1}})(\prod_{k=j+2}^{m} d(a_k|a_{k-1})e(f_k|e_{a_k}))]
\end{aligned}
$$

Then we have

$$P(a_j = k|f, e) = \alpha[j,k] \times \beta[j,k]$$

Question 3(b)

$$P(a_j = k, a_{j+1} = k'|f, e) = \alpha[j,k] \times d(k'|k) \times t(f_{j+1}|e_{k'}) \times \beta[j+1, k']$$

Question 3(c)

Define $\overline{count}(i, a \to a'; \underline{\theta})$ to be the expected number of times the transition $a \to a'$ is seen in the training example $f^{(i)}, e^{(i)}$, for parameters $\underline{\theta}$. Then

$$\overline{count}(i, a \to a'; \underline{\theta}) = \sum_{j=1}^{m-1} p(a_j = a, a_{j+1} = a'|f^{(i)}, e^{(i)}; \underline{\theta})$$

Define $\overline{count}(i, e \rightsquigarrow f; \underline{\theta})$ to be the expected number of times the English word e is paired with the French word f in the training example $f^{(i)}, e^{(i)}$, for parameters $\underline{\theta}$. Then

$$\overline{count}(i, e \rightsquigarrow f; \underline{\theta}) = \sum_{j=1}^{m-1} p(a_j = a_e|f^{(i)}, e^{(i)}; \underline{\theta})$$

The EM algorithm:

- Initialization: set initial parameters $\underline{\theta}^0$ to some value
- For $t = 1 \ldots T$:

– Use the forward-backward algorithm to compute all expected counts of the form

$$\overline{count}(i, a \rightarrow a'; \underline{\theta}^{t-1}) \text{ or } \overline{count}(i, e \rightsquigarrow f; \underline{\theta}^{t-1})$$

– Update the parameters based on the expected counts:

$$d^t(a'|a) = \frac{\sum_{i=1}^{n} \overline{count}(i, a \rightarrow a'; \underline{\theta}^{t-1})}{\sum_{i=1}^{n} \sum_{a'} \overline{count}(i, a \rightarrow a'; \underline{\theta}^{t-1})}$$

$$t^t(f|e) = \frac{\sum_{i=1}^{n} \overline{count}(i, e \rightsquigarrow f; \underline{\theta}^{t-1})}{\sum_{i=1}^{n} \sum_{f} \overline{count}(i, e \rightsquigarrow f; \underline{\theta}^{t-1})}$$