50.007 Machine Learning Homework 4

Victoria Yong 1004455

Q1.

Transition Table, $\alpha_{uv} = \frac{count(u;v)}{count(u)}$

u/v	X	Y	Z	STOP
X	0	0.4	0.4	0.2
Y	0.2	0	0.2	0.6
Z	0.4	0.6	0	0
START	0.5	0	0.5	0

Emission Table, $b_u(o) = \frac{count(A \to e)}{count(A)}$

u/o	A	B	C
X	0.4	0.6	0
Y	0.2	0	0.8
Z	0.2	0.6	0.2

Q2.

START, Z, Y ,STOP

From the emission table above, observation b is most likely to have the state X or Z. Observation c is most likely to have the state Y. From the transition table,

$$P(b = X, c = Y) = P(START, X) * P(X, Y) * P(Y, STOP) = 2/4 * 2/5 * 3/5 = 0.12$$

$$P(b = Z, c = Y) = P(START, Z) * P(Z, Y) * P(Y, STOP) = 2/4 * 3/5 * 3/5 = 0.18$$

Hence it is most likely for b to have the state Z, and c to have the state Y.

Q3.

$Layer\ 1$:

$For\ state\ X,$

$$Transition, \quad \alpha_x^{(1)} = \alpha_{START,x}^{(1)}$$
$$= 0.5$$
$$Emission, \quad \beta_x^{(1)} = \sum_v \alpha_{u,v} b_u(b) \beta_v(2)$$
$$= \alpha_{x,x} b_x(b\ ) \beta_x(u) + \alpha_{x,y} b_y(b\ ) \beta_y(u) + \alpha_{x,z} b_z(b\ ) \beta_z(u)$$
$$= 0 + (0.4)(0.6)(\alpha_{y,STOP}) b_y(c) + 0$$
$$= (0.24)(0.6)(0.8)$$
$$= 0.1152$$

$For\ state\ Y,$

$$\alpha_y^{(1)} = \alpha_{START,y}^{(1)}$$
$$= 0.0$$
$$\beta_y^{(1)} = \sum_v \alpha_{u,v} b_u(b) \beta_v(2)$$
$$= 0.0$$

$For\ state\ Z,$

$$\alpha_z^{(1)} = \alpha_{START,z}^{(1)}$$
$$= 0.5$$
$$\beta_x^{(1)} = \sum_v \alpha_{u,v} b_u(b) \beta_v(2)$$
$$= \alpha_{z,x} b_z(b\ ) \beta_x(u) + \alpha_{z,y} b_y(b\ ) \beta_y(u) + \alpha_{z,z} b_z(b\ ) \beta_z(u)$$
$$= (0.4)(0.6)(a_{x,STOP}) b_x(c) + (0.6)^3(0.8) + 0$$
$$= 0 + 0.1728$$
$$= 0.1728$$

$$\therefore\ s_1 = Z$$

$Layer\ 2$:

$For\ state\ X,$

$$\beta_x^{(2)} = \alpha_{x,STOP}^{(2)} \mathrm{b}_x(\mathrm{c})$$
$$= 0$$
$$\alpha_x^{(2)} = \sum_v \alpha_{vu} \alpha_v(1) b_v(b)$$
$$= 0 + 0 + 0.5(0.4)(0.6)$$
$$= 0.12$$

$For\ state\ Y,$

$$\beta_y^{(2)} = \alpha_{y,STOP}^{(2)} \mathrm{b}_y(\mathrm{c})$$
$$= (0.6)(0.8)$$
$$= 0.48$$

$$\alpha_y^{(2)} = \sum_v \alpha_{vu}\alpha_v(1)b_v(b)$$
$$= 0.5(0.4)(0.6) + 0 + 0.5(0.6)(0.6)$$
$$= 0.3$$

For state Z,

$$\beta_z^{(2)} = \alpha_{z,STOP}^{(2)}\mathrm{b}_z(\mathrm{c})$$
$$= 0$$
$$\alpha_z^{(2)} = \sum_v \alpha_{vu}\alpha_v(1)b_v(b)$$
$$= 0.5(0.4)(0.6) + 0 + 0$$
$$= 0.12$$

Hence,

$$\alpha_x^{(2)}\beta_x^{(2)} = 0$$

$$\alpha_y^{(2)}\beta_y^{(2)} = 0.3(0.48) = 0.144$$

$$\alpha_z^{(2)}\beta_z^{(2)} = 0$$

$$\therefore\ s_2 = Y$$

$$s_1 = Z, \qquad s_2 = Y$$

Q4.

$$\text{Number of states } = |T| + 2$$

$$\begin{aligned}
P(z_i = u|(x,y);\theta) &= \frac{p(x_1, y_1, \dots, x_{i-1}, y_{i-1}, x_i = u, y_i, \dots x_n, y_n; \theta)}{p((x,y);\theta)} \\
&= p(x_1, y_1, \dots x_{i-1}, y_{i-1}) p(x_i y_i, \dots x_n y_n | z_i = u; \theta) \\
&= [p(x_1, \dots, x_{i-1}|z_i = u;\theta) p(y_1, \dots, y_{i-1}|x;\theta)][p(x_i, \dots, x_n|z_i = u;\theta) p(y_i, \dots, y_n|x;\theta)]
\end{aligned}$$

$$\alpha_u(i) = [p(x_1, \dots, x_{i-1}|z_i = u;\theta) p(y_1, \dots, y_{i-1}|x;\theta)]$$
$$\beta_u(i) = [p(x_i, \dots, x_n|z_i = u;\theta) p(y_i, \dots, y_n|x;\theta)]$$

$Forward\ Step$:

$$Since\ \alpha_u(i) = [p(x_1, \dots, x_{i-1}|z_i = u;\theta) p(y_1, \dots, y_{i-1}|x;\theta)]$$

$$\alpha_u(1) = \alpha_{START,u}$$

$For\ i = 1, \dots, n-1$

$$\alpha_u(i+1) = \left[\Sigma_v \alpha_v(i) x_{v,u} b_v(x_i)\right] p(y_i|x_i)$$

$Backward\ Step$:

$$Since\ \beta_u(i) = [p(x_i, \dots, x_n|z_i = u;\theta) p(y_i, \dots, y_n|x;\theta)]$$

$$\beta_u(n) = \alpha_{u,STOP} b_u(x_n) p(y_n|x_n)$$

$For\ i = n-1, \dots, 1$

$$\beta_u(i) = \left[\Sigma_v \alpha_{u,v} b(x_i) \beta_v(i+1)\right] p(y_i|x_i)$$

Complexity of algorithm:

The overall time complexity is $O(n|T|^2)$.

Each sentence has length n, and each word in the sentence has |T| possible states. Computing the forward and backwards algorithm involves |T| steps.