NLP

HW2.

1) $P(\text{Tom}|\text{am}) = \frac{\text{Count}(\text{am, Tom})}{\text{Count}(\text{am})}$

$= \frac{\text{Total count of 'am Tom'}}{\text{Total count of 'am's}}$

$= 3/4 \approx 0.75$

2) Add smoothing to bi-gram model, $P(\text{Tom}|\text{am})$?

$P(\text{Tom}|\text{am}) = \frac{C_i + 1}{N+V} = \frac{C(w_{n-1} w_n) + 1}{C(w_{n-1}) + V}$

$= \frac{C(\text{am}|\text{Tom}) + 1}{C(\text{am}) + \text{Vocabulary}}$

$= \frac{4 (=3+1)}{4 + 12}$ (12 unique words)

$= \frac{4}{16} = \frac{1}{4} = 0.25$

(c) Linear Interpolation smoothing b/n MLE and mix-likelihood region.

$\lambda_1 = 1/2 \qquad \lambda_2 = 1/2$.

$$P(\text{Tom}|\text{am}) = \lambda_1 P(\text{Tom}) + \lambda_2 P(\text{Tom}|\text{am})$$

$$= \frac{1}{2}\left(\frac{7}{42}\right) + \frac{1}{2}\left(\frac{3}{4}\right)$$

$$= \frac{1}{12} + \frac{3}{8} = \frac{44}{12(8)} \quad \frac{11}{24}$$

$$= 11/24$$

$$\cong 0.4583$$

(d) probability of <S> I like Jane </S>

$$P(\text{<S> I like Jane </S>}) = P(\text{I}|\text{<S>}) \times P(\text{like}|\text{I}) \times P(\text{Jane}|\text{like}) \times P(\text{</S>}|\text{Jane})$$

$$= \frac{\text{Count(<S> I)}}{\text{Count(<S>)}} \times \frac{\text{Count(I like)}}{\text{Count(I)}} \times \frac{\text{Count(like Jane)}}{\text{Count(like)}} \times \frac{\text{Count(Jane </S>)}}{\text{Count(Jane)}}$$

$$= \frac{5}{7} \times \frac{2}{63} \times \frac{1}{3} \times \frac{2}{2} = \frac{5}{63}$$

$$\approx 0.0794$$

(e) <s> I like Jane </s> by using Smoothing.

$$P(\text{<s> I like Jane </s>}) = P(\text{I} \mid \text{<s>})\, P(\text{like} \mid \text{I})\, P(\text{Jane} \mid \text{like})\, P(\text{</s>} \mid \text{Jane})$$

$$= \frac{C(\text{<s> I}) + 1}{C(\text{<s>}) + V} \times \frac{C(\text{I like}) + 1}{C(\text{I}) + V} \times \frac{C(\text{like Jane}) + 1}{C(\text{like}) + V} \times \frac{C(\text{Jane </s>}) + 1}{C(\text{Jane}) + V}$$

$$= \frac{5+1}{7+12} \times \frac{2+1}{6*12} \times \frac{1+1}{8+12} \times \frac{2+1}{2+12}$$

$$= \frac{6}{19} \times \frac{3}{18} \times \frac{2}{15} \times \frac{3}{14}$$

$$= 0.0015037$$

(f) "<s> I like Jane </s>" with interpolation smoothing

$$\lambda_1 = \frac{1}{2} \qquad \lambda_2 = \frac{1}{2}$$

$$P(\langle s\rangle\ I \text{ like Jane } \langle /s\rangle) = \left[\lambda_1 P(I) + \lambda_2 P\left(\frac{I}{\langle s\rangle}\right)\right] \cdot \left[\lambda_1 P(\text{like}) + \lambda_2 P\left(\frac{\text{like}}{I}\right)\right] \cdot \left[\lambda_1 P(\langle /s\rangle) + \lambda_2 P\left(\frac{\langle /s\rangle}{\text{Jane}}\right)\right]$$

$$= \left[\frac{1}{2}\cdot\frac{6}{42} + \frac{1}{2}\left(\frac{5}{7}\right)\right] + \left[\frac{1}{2}\left(\frac{3}{42}\right) + \frac{1}{2}\left(\frac{2}{6}\right)\right] + \left[\frac{1}{2}\frac{2}{42} + \frac{1}{2}\frac{1}{3}\right] + \left[\frac{1}{2}\frac{7}{42} + \frac{1}{2}\frac{2}{2}\right]$$

$$= \frac{1}{2}\ \frac{6}{7}\ \left[\frac{1}{2}\left(\frac{17}{42}\right)\right]\ \left[\frac{1}{2}\ \frac{16}{42}\right]\ \left[\frac{1}{2}\ \frac{49}{42}\right]$$

$$= \frac{3}{7}\ \frac{17}{84}\ \frac{4}{21}\ \frac{7}{21}$$

$$= \frac{17}{1764}.$$