< Chapter 3 >

3.1. 1) Trigram Probability : $P(w_1 w_2 \cdots w_n) \approx \prod_{k=1}^{n} P(w_k \mid w_{k-2:k-1})$

$$= \prod_{k=1}^{n} \frac{C(w_{k-2}w_{k-1}w_k)}{C(w_{k-2}w_{k-1}w)} \approx \prod_{k=1}^{n} \frac{C(w_{k-2}w_{k-1}w_k)}{C(w_{k-2}w_{k-1})}$$

2) <s> <s> I am Sam </s>
<s><s> Sam I am </s>
<s><s> I do not like green eggs and ham </s>

$$P(\langle s\rangle\langle s\rangle \text{ I am Sam } \langle /s\rangle) = P(I \mid \langle s\rangle\langle s\rangle)\cdot P(am \mid \langle s\rangle I)\cdot P(Sam \mid I\ am)\cdot P(\langle /s\rangle \mid am\ Sam)$$

$$= \frac{2}{3}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot 1$$

$$= \frac{1}{6}$$

3.2 1) $P(\langle s\rangle \text{ I want chinese food } \langle /s\rangle) = P(I \mid \langle s\rangle)\, P(want \mid I)\, P(chinese \mid want)\, P(food \mid chinese)\, P(\langle /s\rangle \mid food)$

$$= 0.25 \times 0.33 \times 0.0065 \times 0.52 \times 0.68 = 1.9 \times 10^{-4}$$

2) Laplace Smoothing

$$P(I \mid \langle s\rangle)\, P(want \mid I)\, P(chinese \mid want)\, P(food \mid chinese)\, P(\langle /s\rangle \mid food)$$

$$= \frac{C(\langle s\rangle I)+1}{\sum_w (C(\langle s\rangle w)+1)} \cdot \frac{C(\text{want chinese})+1}{\sum_w (C(\text{want } w)+1)} \cdots \frac{C(\text{food } \langle /s\rangle)+1}{\sum_w (C(\text{food } w)+1)}$$

$$= \frac{C(\langle s\rangle I)+1}{C(\langle s\rangle)+|W|} \cdot \frac{C(\text{want chinese})+1}{C(\text{want})+|W|} \cdots \frac{C(\text{food } \langle /s\rangle)+1}{C(\text{food})+|W|} = 0.19 \times 0.21 \times 0.0029 \times 0.052 \times 0.40 = 2.4 \times 10^{-6}$$

3.3 The unsmoothed probability is higher

→ because as the zero-probability turns to non-zero during smoothing, a mass number of previously-nonzero-probability moves to previously-zero-probability.

3.4. $P(Sam \mid am) = \frac{C(\text{am Sam})+1}{\sum_w (C(\text{am } w)+1)} = \frac{C(\text{am sam})+1}{C(\text{am})+|W|} = \frac{3}{3+11} = 0.214$

$W = \{I, am, Sam, do, not, like, green, eggs, and, \langle s\rangle, \langle /s\rangle\} \Rightarrow |W| = 11$

3.5 The p of four possible sentences over $\{a, b\}$ are

① $P(\langle s\rangle\, a\, a) = P(a \mid \langle s\rangle)\, P(a \mid a) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

② $P(\langle s\rangle\, a\, b) = P(a \mid \langle s\rangle)\, P(b \mid a) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

③ $P(\langle s\rangle\, b\, b) = P(b \mid \langle s\rangle)\, P(b \mid b) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

④ $P(\langle s\rangle\, b\, a) = P(b \mid \langle s\rangle)\, P(a \mid b) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

< Confusing Point >

> $P(a|a), P(b|a), P(a|b), P(b|b)$ Counting
>
> $\Rightarrow P(a|a) = \frac{C(a\,a)}{\sum_w C(a\,w)}$
>
> → Since there's no termination pseudoword (</s>), the count becomes 2, instead of 4

↝ Therefore, the sum of probabilities of the possible sentences over $\{a, b\}$ is 1.0

The sum of p of all possible 3 word sentences over $\{a, b\}$ is,

$$P(\langle s\rangle\, a\, a\, a) + P(\langle s\rangle\, a\, a\, b) + \cdots + P(\langle s\rangle\, b\, a\, a) + P(\langle s\rangle\, b\, a\, b)$$

$$= P(\langle s\rangle\, a\, a)\underbrace{(P(a|a)+P(b|a))}_{=1} + P(\langle s\rangle\, a\, b)\underbrace{(P(a|b)+P(b|b))}_{=1} + P(\langle s\rangle\, b\, b)\underbrace{(P(a|b)+P(b|b))}_{=1} + P(\langle s\rangle\, b\, a)\underbrace{(P(a|a)+P(b|a))}_{=1} = 1.0$$

⇒ The sum of probabilities for 2-word sentences and 3-word sentences > 1.0

∴ the bigram model doesn't assign a single probability distribution

3.6 $P(w_3|w_1w_2) = \frac{C(w_1w_2w_3)+1}{\sum_w(C(w_1w_2w)+1)} = \frac{C(w_1w_2w_3)+1}{C(w_1w_2)+V}$

3.8, 3.10 : github