

Språkmodeller

n -grammodeller

Brukes for sannsynlighetsberegning

Minikorpus:

<s> Lise elsker å danse <\s>

<s> Ola hater å sykle <\s>

<s> Peter liker å danse <\s>

Unigrammer ($n=1$): <s>, <Lise>, <elsker>, <danse>, <\s>

Bigrammer ($n=2$): <<s>, Lise>, <Lise, elsker>, <elsker, å>, <å, danse>, <danse, <\s> >

$$P(< s > \text{ Ola hater å danse } < \backslash s >)$$

Minikorpus:

<s> Lise elsker å danse <\s>

<s> Ola hater å sykle <\s>

<s> Peter liker å danse <\s>

$$P(\text{Ola} | < s >) * P(\text{hater} | \text{Ola}) * P(\text{å} | \text{hater}) * P(\text{danse} | \text{å}) * P(< \backslash s > | \text{danse})$$

$$= \frac{\text{count}(< s >, \text{Ola})}{\text{count}(< s >)} * \frac{\text{count}(\text{Ola}, \text{hater})}{\text{count}(\text{Ola})} * \frac{\text{count}(\text{hater}, \text{å})}{\text{count}(\text{hater})} * \frac{\text{count}(\text{å}, \text{danse})}{\text{count}(\text{å})} * \frac{\text{count}(\text{danse}, < \backslash s >)}{\text{count}(\text{danse})}$$

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

$$= 0.33 * 1 * 1 * 0.5 * 1 = 0.165$$

= 16.5% sjanse for setningen