

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> >

$P(\langle s \rangle \text{ Ola hater å danse } \langle \backslash s \rangle)$

Minikorpus:

$\langle s \rangle$ Lise elsker å danse $\langle \backslash s \rangle$

$\langle s \rangle$ Ola hater å sykle $\langle \backslash s \rangle$

$\langle s \rangle$ Peter liker å danse $\langle \backslash s \rangle$

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

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

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

$$= 0.33 * 1 * 1 * 0.66 * 1 = 0.217$$

= 21.7% sjans for setningen