IR 26.5.08

Notiztitel 26.05.2008

$P(a|b) = P(b|a) \cdot P(a) / P(b)$

$$O(R|q_k,\vec{x}) = \frac{P(R|q_k,\vec{x})}{P(\bar{R}|q_k,\vec{x})} = \frac{P(\vec{x}|R,q_k)\cdot P(R|q_k)\cdot \cancel{P(\vec{x}|q_k)}}{P(\vec{x}|\bar{R},q_k)\cdot P(\bar{R}|q_k)\cdot \cancel{P(\vec{x}|q_k)}}$$

Generalität von q_k

$$\log \frac{1-q_{ik}}{q_{ik}} \approx \log \frac{1-\frac{n_i}{N}}{n_i/N} = \log \frac{\cancel{N}\left(\frac{N-n_i}{\cancel{N}}\right)}{n_i} = \log \frac{N-n_i}{n_i}$$

$$\approx \log \frac{N}{n_i}$$

IDF-Gewicht

IDF-Formel:

$$\sum_{t_i \in q_k \cap d_m} \log \frac{N-n_i}{n_i}$$

$p_1 = P(t_1 \text{ kommt in rel. Dok. vor}) = \frac{r_1}{r} = \frac{8}{12} = \frac{2}{3}$

$q_1 = P(t_1 \text{ kommt in irrel. Dok. vor}) = \frac{n_1 - r_1}{N - r} = \frac{3}{8}$

$p_2 = \frac{7}{12} \quad q_2 = \frac{4}{8} = \frac{1}{2} \quad O(R|q_2) = \frac{r}{N-r} = \frac{12}{8} = \frac{3}{2}$

$$O(R|(1,1)) = O(R|q_2) \cdot \frac{p_1}{q_1} \cdot \frac{p_2}{q_2} = \frac{3}{2} \cdot \frac{2 \cdot 7 \cdot 8 \cdot 2}{3 \cdot 12 \cdot 3 \cdot 1} = \frac{28}{9}$$

$$P(R|(1,1)) = \frac{\theta}{1+\theta} = \frac{28}{9\left(1+\frac{28}{9}\right)} = \frac{28 \cdot 9}{9 \cdot 37} = \frac{28}{37} \approx 0,76$$

$$O(R|(1,0)) = O(R|q_2) \cdot \frac{p_1}{q_1} \cdot \frac{1-p_2}{1-q_2} = \frac{3}{2} \cdot \frac{2 \cdot 8 \cdot 5 \cdot 2}{3 \cdot 3 \cdot 12 \cdot 1} = \frac{20}{9}$$

$$P(R|(1,0)) = \frac{\theta}{1+\theta} = \frac{20}{9\left(1+\frac{20}{9}\right)} = \frac{20}{29} \approx 0,69$$

$P(t_1 t_2 | R) = P(t_1|R) \cdot P(t_2|R) = p_1 \cdot p_2 = \frac{2}{3} \cdot \frac{7}{12} = \frac{7}{18}$ vs. $\frac{4}{12} = \frac{1}{3}$