

IR 18.5.06

Notiztitel

18.05.2006

$$p_1 = P(x_1 \text{ kommt in rel. Dok. vor}) = \frac{8}{12} = \frac{2}{3}$$

$$q_1 = P(x_1 \text{ kommt in i-rel. Dok. vor}) = \frac{3}{8}$$

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

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

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

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

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

$$P(x_1 x_2 | R) = p_1 \cdot p_2 = \frac{2}{3} \cdot \frac{7}{12} = \frac{7}{18} \quad \text{vs.} \quad \frac{4}{12} = \frac{1}{3}$$

Kosten für Retrieval

$$\text{Sei } C=1, \bar{C}=2$$

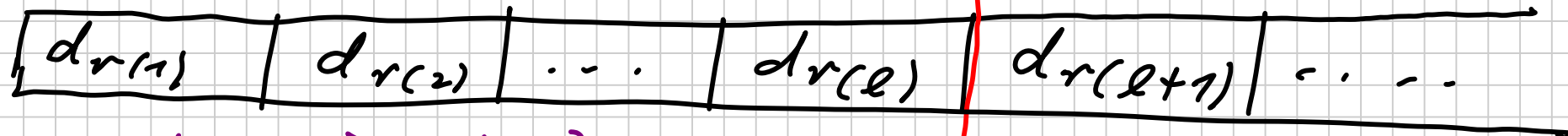
$$P(R|q,d) = \underline{0,7}$$

$$\begin{aligned} EC(q,d) &= C \cdot P(R|q,d) + \bar{C} \cdot [1 - P(R|q,d)] \\ &= 1 \cdot 0,7 + 2 \cdot 0,3 = \underline{1,3} \end{aligned}$$

$$P(R|q,d') = \underline{0,5}$$

$$EC(q,d') = 1 \cdot 0,5 + 2 \cdot 0,5 = \underline{1,5}$$

minimiere Kosten $\sum_{i=1}^l EC(q, d_{r(i)})$



$$EC(q, d_{r(1)}) = EC(q, d_{r(2)}) \leq \dots \leq EC(q, d_{r(l)}) \leq EC(q, d_{r(l+1)})$$

minimiere Summe dieser Kosten

→ ordne nach aufsteigenden Kosten!

$$E C(q, d) \leq E C(q, d') \Leftrightarrow$$

$$c \cdot P(R|q, d) + \bar{c} [1 - P(R|q, d)] \leq c P(R|q, d') + \bar{c} [1 - P(R|q, d)']$$

$$c \cdot P + \bar{c} - \bar{c} \cdot P \leq c \cdot P' + \bar{c} - \bar{c} \cdot P' \quad | \cdot (-1)$$

$$(\bar{c} - c) P \geq (\bar{c} - c) P' \quad | : (\bar{c} - c)$$

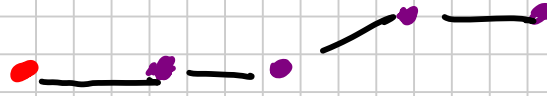
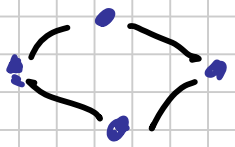
$$P \geq P'$$

Ordnung nach steigenden Kosten

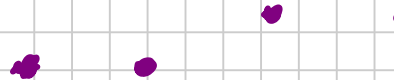
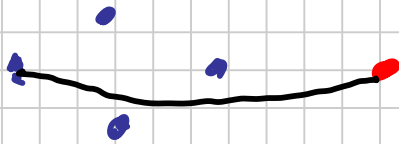
$\hat{=}$ Ordnung nach fallender Relevanzwahrsch.

(für $c < \bar{c}$)

Clustering



single link



complete link



average link