

Ralf Gehrer

Fabian Lindenberg

Distributed

Co-Clustering

Szenario

| | Autor | Verlag | Produktion | Darsteller | Regie |
|-----------------|-------|--------|------------|------------|-------|
| Der Gefangene | 1 | 1 | 0 | 0 | 0 |
| Sakrileg | 1 | 1 | 0 | 0 | 0 |
| L.A. Crash | 0 | 0 | 1 | 1 | 1 |
| Minority Report | 0 | 0 | 1 | 1 | 1 |

Szenario

| | Autor | Produktion | Verlag | Darsteller | Regie |
|-----------------|-------|------------|--------|------------|-------|
| Der Gefangene | 1 | 0 | 1 | 0 | 0 |
| L.A. Crash | 0 | 1 | 0 | 1 | 1 |
| Sakrileg | 1 | 0 | 1 | 0 | 0 |
| Minority Report | 0 | 1 | 0 | 1 | 1 |

Szenario

| | 1 | 1 | 1 | 2 | 2 |
|---|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0 | 0 |
| 1 | 0 | 1 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 | 0 | 0 |
| 2 | 0 | 1 | 0 | 1 | 1 |

Szenario

| | 1 | 1 | 1 | 2 | 2 |
|---|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0 | 0 |
| 1 | 1 | 0 | 1 | 0 | 0 |
| 2 | 0 | 1 | 0 | 1 | 1 |
| 2 | 0 | 1 | 0 | 1 | 1 |

Szenario

| | 1 | 1 | 2 | 2 | 2 |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1 | 1 | 1 |
| 2 | 0 | 0 | 1 | 1 | 1 |

Algorithmus

repeat

for each row $i = 1..m$ **do**

M **for each** row group label = $1..k$ **do**
if error is smaller
 $r(i) = p$

R reorganize rows according to new labels

M **for each** column $j = 1..n$ **do**
 ...

R reorganize columns

until error does not decrease

Kostenfunktion

$$\begin{aligned} \sum_{j=1}^{\ell} \sum_{u=0}^1 n_u(x^j) \log \frac{1}{P_{\Psi^{t+1}(x),j}^t(u)} \\ \leq \sum_{j=1}^{\ell} \sum_{u=0}^1 n_u(x^j) \log \frac{1}{P_{i,j}^t(u)} \end{aligned}$$

Kostenfunktion

$$T(D; k, \ell, \Psi, \Phi) :=$$

$$\begin{aligned} & \log^* k + \log^* \ell + \sum_{i=1}^{k-1} \lceil \log \bar{a}_i \rceil + \sum_{j=1}^{\ell-1} \lceil \log \bar{b}_j \rceil \\ & + \sum_{i=1}^k \sum_{j=1}^{\ell} \lceil \log(a_i b_j + 1) \rceil + \sum_{i=1}^k \sum_{j=1}^{\ell} C(D_{i,j}) \end{aligned}$$

Kostenfunktion

$$C(A) := \sum_{i=0}^1 n_i(A) \log \left(\frac{n(A)}{n_i(A)} \right) = n(A) H(P_A(0))$$

Distributed

Co-Clustering

S. Papadimitriou, J. Sun

“DisCo: Distributed Co-clustering
with Map-Reduce”

D. Chakrabarti, S. Papadimitriou,
D. Modha, C. Faloutsos

“Fully Automatic Cross-
Associations”