7.1 Ü IR 20.11.18

a) R

b) TAAT (score berechnen) $N = 4$, Top-k

mindestens N speicher

Addition oder Multiplikation

DAAT

$k=2$

$d_1 : S_1$
$d_2 : S_2$ → $d_1 : S_1$, $d_3 : S_3$ → $\boxed{d_1 : S_2 \quad d_3 : S_3}$ $k+1$ speicher

$d_3 : S_3$ $d_4 : S_4$

~~ger~~ kleiner als S_1 S_2
d_3 weg schmeißen
sonst kleinst von d_1 oder d_2
wegschmeißen

TAAT mehr Speicher als DAAT

R

c) ~~NO~~NRA
F
nicht immer alle Posting-list listen muss

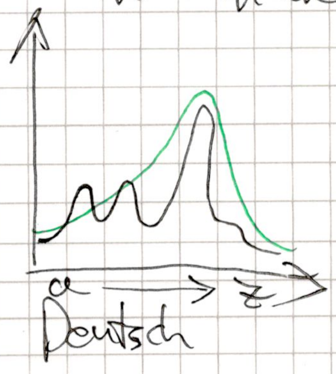
d) R

e) F

Sampling von $m \ll N$ Dokumenten

→ Verteilung Anfangsbuchstaben

→ ~~hoch~~ hoch rechnen



~~08.12.18~~

~~V CBM~~
Ü IR

10.1

a) F increase

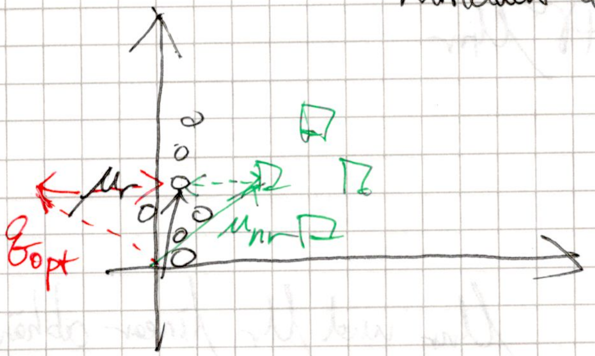
q_0 ... Query Vektor

P_r ... Relevante Doc

P_{nr} ... Nicht relevante Doc

b) R $q_{opt} = \underbrace{\frac{1}{|D_r|}\sum_{d_j \in D_r} d_j}_{\mu_r} + \Big[\underbrace{\frac{1}{|D_r|}\sum_{d_j \in D_r} d_j}_{\mu_r} - \underbrace{\frac{1}{|D_{nr}|}\sum_{d_j \in D_{nr}} d_j}_{\mu_{nr}}\Big]$

Mittelwert d. relevanten Doc



c) F

d) ~~Richtig~~

e) R

f) F kann automatisch erstellt werden

welche Term häufig zusammen

Ähnlichkeit beobachten

10.2

$q_m = \alpha q_0 + \beta \mu_r - \gamma \mu_{nr}$

ursprüngliche Query

① Query und Dok-Vektor bestimmen

② μ_r und μ_{nr} bestimmen

③ q_m bestimmen

Term	q_0	d_1	d_2	d_3	μ_r	μ_{nr}	q_m
tasty	1	0	0	2	0	2	0,5
hot	1	1	3	2	2	2	2
chocolate	1	2	2	0	2	0	2,5
tea	0	0	0	2	0	2	0
milk	0	2	2	1	2	1	1,25

$\mu_r = \frac{1}{|D_r|}\sum_{d_j \in D_r} d_j$

$\mu_{nr} = \frac{1}{|D_{nr}|}\sum_{d_j \in D_{nr}} d_j$

$q_m[\text{'tasty'}] = \alpha \cdot 1 + \beta \cdot 0 - \gamma \cdot 2$

$= 1 + 0,75 \cdot 0 - 0,25 \cdot 2$

$= 0,5$

Achtung: keine negative Wert im q_m → Wert auf 0 gesetzt

14.1 a) F Intranet wie z.B. random Server

08.01.21
Ü IR

b) F andere 2 subclasses Transactional needs

c) R hashed ~~the~~ die Kurzzusammenfassung min-hash

d) F laut Graphtechnische Grundlagen (Deadlinks nicht mitgezählt)

e) F

f) F

g) R

h) R

14.2

$$P = \begin{bmatrix} P_{11} & \cdots & P_{1j} \\ \vdots & \ddots & \vdots \\ P_{i1} & \cdots & P_{ij} \end{bmatrix}$$

In meisten Fällen: $i=j$

wenn Sackgasse $(o(i)=0)$:

$P_{ij} = \frac{1}{n}$

Sonst.

1. $(i,j) \notin E$: $P_{ij} = \frac{\alpha}{n}$
2. $(i,j) \in E$: $P_{ij} = \frac{\alpha}{n} + \frac{1-\alpha}{o(i)}$ — Wsk nicht teleportiert ist.

$\alpha = 0$: Teleportationsmöglichkeit

$$P = \begin{array}{c|ccc} & A & B & C \\ \hline A & 0 & 1 & 0 \\ B & \frac{1}{2} & 0 & \frac{1}{2} \\ C & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}$$

→ ~~gar~~ keine Möglichkeit

weil kein A → A

$\frac{0}{3} + \frac{1-0}{2}$

zwei Out-degree $\Rightarrow o(i)=2$

Sackgasse

3 Unterscheidungsmerkmale

- Sackgasse
- $(i,j) \notin E$
- $(i,j) \in E$

$\alpha = \frac{1}{2}$:

$$P = \begin{array}{c|ccc} & A & B & C \\ \hline A & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ B & \frac{5}{12} & \frac{1}{6} & \frac{5}{12} \\ C & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}$$

→ geringere Möglichkeit

1. $\frac{1}{2}/3 = \frac{1}{6}$

2. $\frac{\frac{1}{2}}{3} + \frac{1-\frac{1}{2}}{o(i)=1}$

⇓

1 Ausgangskante

$\vec{A_k} = \vec{X_k} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ — hier immer 0 da $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \vdots \end{pmatrix}$

— höchstens auf 1 normalisiert $\Rightarrow$ laut ~~Aufgabestellung~~

— $b = c$

$$\vec{H_k} = \begin{pmatrix} \overset{b}{1} + \overset{c}{1} = 2 \\ \overset{a}{0} + \overset{c}{1} = 1 \\ \overset{a}{0} + \overset{b}{1} = 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \xrightarrow{\text{normalisieren}} \begin{pmatrix} 1 \\ 1/2 \\ 1/2 \end{pmatrix}$$

14.4 a) ~~$\forall_{i,j} : p_{ij} \geq \frac{\alpha}{n}$~~

$\forall_{x_i} : x_i \geq \frac{\alpha}{n} \quad x_i \in \vec{X_k}$

$\vec{X_k} = P^T \cdot \vec{X_{k-1}}$

NB: $1 \leq i, j \leq n$

$0 \leq \alpha \leq 1$

1. $\forall_{i,j} : p_{ij} \geq \frac{\alpha}{n} \quad p_{ij} \in P$

- Sackgasse $(O(i) = 0)$: $\frac{1}{n} \geq \frac{\alpha}{n}$
- $(i,j) \notin E$: $\frac{\alpha}{n} \geq \frac{\alpha}{n}$
- $(i,j) \in E$: $\frac{\alpha}{n} + \frac{1-\alpha}{O(i)} \geq \frac{\alpha}{n}$

2. $\vec{X_{k,i}} = \vec{X_{k-1,i}} \cdot \vec{z_i}$ $\quad \vec{z_i}$: i-te Zeile aus P^T — Jede Element aus $\vec{z_i}$ ist größer ~~als~~ gleich als $\frac{\alpha}{n}$

$\vec{z_i} = \vec{u_i} + \vec{v_i}$ mit $\vec{u_i} = \begin{pmatrix} \frac{\alpha}{n} \\ \frac{\alpha}{n} \\ \frac{\alpha}{n} \\ \vdots \\ \frac{\alpha}{n} \end{pmatrix}$ jede Elemente von $\vec{u_i}$ ist $\frac{\alpha}{n}$

$\vec{V_i} = \begin{pmatrix} \vec{z_{i,1}} - \frac{\alpha}{n} \\ \vdots \\ \vec{z_{ij}} - \frac{\alpha}{n} \end{pmatrix} \begin{matrix} \geq 0 \\ \geq 0 \\ \vdots \\ \geq 0 \end{matrix} \Rightarrow \vec{V_{ij}} \geq 0$

$$\vec{X_{k,i}} = \vec{X_{k-1,i}} \cdot (\vec{u_i} + \vec{v_i}) = \underbrace{\vec{X_{k-1,i}} \cdot \vec{u_i}}_{1} + \underbrace{\vec{X_{k-1,i}} \cdot \vec{v_i}}_{2}$$

1. $\vec{X_{k-1,i}} \cdot \vec{u_i} = \underbrace{\sum_{i=1}^{n} \vec{X_{k-1,i}}}_{=1} \cdot \frac{\alpha}{n} = \frac{\alpha}{n}$

2. $\vec{X_{k-1,i}} \cdot \vec{V_i} \geq 0$

$\Rightarrow \vec{X_{k,i}} \geq \frac{\alpha}{n}$

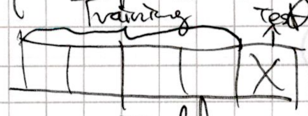
19...19.1 Ü IR 22.01.19 29.01.19

a) ~~R~~ muss kein Doku sein

b) accuracy = 1-error R
Fehler ~~def~~ Def: $y_i \,!= \hat{y}_i$ Summe aller

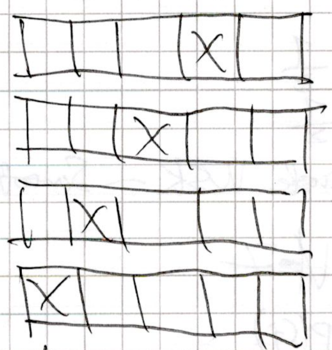
c) $Y = \{(d_i, l_i) \mid 0 \le i \le n\}$ F
Label ~~fo~~ Vorgegeben

d) F ~~Training Daten nicht classifier~~

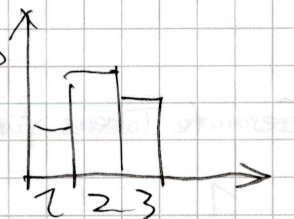
e) 

5 times F ~~Klausurrelevant~~

5-Fold

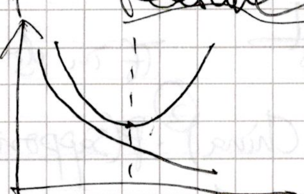


f) R

g) 

R

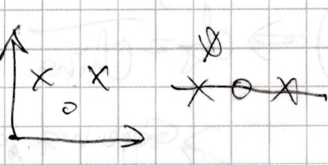
h) F ~~Feature Selection~~ ~~B~~ kann in beide Richtung



i) R

j) F
x
o ~~x~~ ~~übernächste~~ nächste 1NN → x die nächste ungerade Zahl
2NN → u.d
Δ ~~nächste~~ übernächste 3NN → x nicht Centroid

k) R

l) F 

$P(China) = P(\overline{China}) = \frac{1}{2}$

~~$\hat{P}(t|C_j) = \frac{df_{C_j}(t)+1}{g_j+|G_j|}$~~

$t = \{Tainan, Sapporo\}$ $\quad \hat{P}(t|C_j) = \frac{df_{C_j}(t)+1}{g_j+|G_j|} = \frac{df_{C_j}(t)+1}{g_j+g_j}$ $\quad$ G – Menge, g – Zahlen

$t_k = \{Taipei, Macao, Japan, Shanghai, Osaka\}$ $\quad \boxed{\to g_j+2}$ In Vorlesungsfolien stets 2 ??? deshalb 2 benutzen

$$\hat{P}(\overline{t_k}|C_j) = 1 - \hat{P}(t_k|C_j) = \frac{g_j - df_{C_j}(t_k)+1}{g_j+g_j}$$

$\boxed{\to g_j+2}$ In Vorlesungsfolien stets 2 ??? deshalb 2 benutzen

$\hat{P}(Tainan|China) = \frac{2+1}{2+2} = \frac{3}{4}$

$\hat{P}(Sapporo|China) = \frac{0+1}{2+2} = \frac{1}{4}$

~~$\hat{P}(Tainan|\overline{China}) = \frac{1+1}{2+2} = \frac{1}{2}$~~

~~$\hat{P}(Sapporo|\overline{China})$~~

~~$\hat{P}(Japan|\overline{China})$~~

$\hat{P}(\overline{Japan}|China) = \hat{P}(\overline{Osaka}|China) = \frac{2-0+1}{2+2} = \frac{3}{4}$

$\hat{P}(\overline{Taipei}|China) = \frac{2-1+1}{2+2} = \frac{1}{2} = \hat{P}(\overline{Macao}|China)$

$\hat{P}(China|d_5) = \frac{1}{2}\cdot\left(\frac{3}{4}\cdot\frac{1}{4}\right)\cdot\left(\frac{3}{4}\cdot\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\right) = \hat{P}(\overline{Shanghai}|China) = \frac{27}{4096}$

$\hat{P}(Tainan|\overline{China}) = \frac{1+1}{2+2} = \frac{1}{2}$

$\hat{P}(Sapporo|\overline{China}) = \frac{2+1}{2+2} = \frac{3}{4}$

$\hat{P}(\overline{Japan}|\overline{China}) = \frac{2-1+1}{2+2} = \frac{1}{2} = \hat{P}(\overline{Osaka}|\overline{China})$

$\hat{P}(\overline{Taipei}|\overline{China}) = \hat{P}(\overline{Macao}|\overline{China}) = \hat{P}(\overline{Shanghai}|\overline{China}) = \frac{2-0+1}{2+2} = \frac{3}{4}$

$\hat{P}(\overline{China}|d_5) = \frac{1}{2}\cdot\left(\frac{1}{2}\cdot\frac{3}{4}\right)\cdot\left(\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{3}{4}\cdot\frac{3}{4}\right) = \frac{81}{4096}$

$\hat{P}(China|d_5) < \hat{P}(\overline{China}|d_5) : d_5 \Rightarrow \overline{China}$

$$\chi^2(t,c) = \sum_{e_t \in \{0,1\}} \sum_{e_c \in \{0,1\}} \frac{N_{e_t e_c} - E_{e_t e_c}}{E_{e_t e_c}}$$ E – Erwartungswert

$$\boxed{E_{e_t e_c} = N \cdot P(t) \cdot P(c)}$$

$$\chi^2(t,c) = \frac{N_{11} - E_{11}}{E_{11}} + \frac{N_{01} - E_{01}}{E_{01}} + \frac{N_{10} - E_{10}}{E_{10}} + \frac{N_{00} - E_{00}}{E_{00}}$$

$$E_{11} = N \cdot \frac{N_{1\cdot}}{N} \cdot \frac{N_{\cdot 1}}{N} = 1{,}76$$

$$E_{01} = N \cdot \frac{N_{0\cdot}}{N} \cdot \frac{N_{\cdot 1}}{N} = \cancel{7{,}76}\ 1150{,}24$$

$$E_{10} = N \cdot \frac{N_{1\cdot}}{N} \cdot \frac{N_{\cdot 0}}{N} = 151{,}24$$

$$E_{00} = N \cdot \frac{N_{0\cdot}}{N} \cdot \frac{N_{\cdot 0}}{N} = 98696{,}76$$

$\chi^2(t,c) = 588{,}7 \Rightarrow$ brazil, roasted

b)

	$\cancel{MI()}$	χ^2
dependent	1	∞
independent	0	0