IR Class Activity 1
Eisha Tir Raazia
17k-3730
Sec: GR1

Date:

Problem no. 01

(a) Cumulative gain of the two queries.

for q_1:

$$\sum_{i=1}^{7} rel_i = 2+0+2+2+1+1+1 = 9.$$

for q_2:

$$\sum_{i=1}^{7} rel_i = 2+2+2+1+1+0+0 = 8$$

(b) Discounted Cumulative gain (DCG)

for q_1:

i	doc	rel_i	$\log_2(i+1)$	$\frac{rel_i}{\log_2(i+1)}$
1	4	2	1	2
2	5	0	1.525	0
3	6	2	2	1
4	7	2	2.322	0.861
5	1	1	2.5849	0.386
6	8	1	2.807	0.356
7	9	1	3	0.333

$$\sum_{i=1}^{7} = 4.9339$$

$\therefore \boxed{DCG_1 = 4.93 \approx 5}$

for q_2

i	doc	rel_i	$\log_2(i+1)$	$\frac{rel_i}{\log_2(i+1)}$
1	2	2	1	2
2	3	2	1.525	1.261
3	8	2	2	1
4	4	1	2.322	0.430
5	5	1	2.5849	0.3868
6	9	0	2.807	0
7	1	0	3	0

$$\sum = 5.0778$$

$\therefore \boxed{DCG_2 = 5.077}$

DALMATIAN

Date:

(c) Normalized DCG:

query # 01

i	rel_i	Ideal rel_i	$\log_2(i+1)$	$\frac{rel_i}{\log_2(i+1)}$	$\frac{\text{Ideal } rel_i}{\log_2(i+1)}$
1	2	2	1	2	2
2	0	2	1.585	0	1.261
3	2	2	2	1	1
4	2	1	2.322	0.861	0.430
5	1	1	2.585	0.386	0.386
6	1	1	2.807	0.356	0.356
7	1	0	3	0.333	0

$DCG = \sum_{i=1}^{7} = 4.936$ $\quad \sum_{i=1}^{7} = 5.433 = IDCG$.

$$\Rightarrow nDCG_7 = \frac{DCG_7}{IDCG_7} = \frac{4.936}{5.433} = 0.908$$ Ans.

query # 02

i	rel_i	Ideal rel_i	$\log_2(i+1)$	$\frac{rel_i}{\log_2(i+1)}$	$\frac{\text{Ideal } rel_i}{\log_2(i+1)}$
1	2	2	1	1	1
2	2	2	1.525	1.261	1.261
3	2	2	2	1	1
4	1	1	2.322	0.430	0.430
5	1	1	2.584	0.386	0.3868
6	0	0	2.807	0	0
7	0	0	3	0	0

$DCG = 5.0778$ $\quad IDCG = 5.0778$

$$nDCG = \frac{DCG}{IDCG} = \frac{5.0778}{5.0778} = 1$$

Date:

Problem no. 02

Yes = 1
No = 0 } Annotated by judges on docs.

Judge 2

		Yes (1)	No (0)	Total
Judge 1	Yes (1)	2	2	4
	No (0)	2	2	4
	Total	4	4	8

(a) Kappa Measure.

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$ → hypothetical probability.

P(A)

$$P(A) = 4/8 = 1/2 = 0.5$$

for P(E)

$$P(NR) = \frac{4+4}{8+8} = \frac{8}{16} = 0.5$$

$$P(Rel) = \frac{4+4}{8+8} = \frac{8}{16} = 0.5.$$

∴ Now, $P(E) = P(NR)^2 + P(Rel)^2 = 0.5^2 + 0.5^2 = 0.5.$

now,

$$Kappa = \frac{P(A) - P(E)}{1 - P(E)} = \frac{0.5 - 0.5}{1 - 0.5} = \frac{0}{0.5} = 0$$ Ans.

ate:

(b) Precision, recall and F1 when ~~judges~~ doc. is relevant only when two judges agree.

* Precision $= \frac{2}{5} = 0.4$

* Recall $= \frac{2}{2} = 1.$

* F1 $= \frac{2(0.4)(1)}{0.4+1} = 0.5714$

Formulae:

Precision $= \frac{\text{Rel. ret}}{\text{Total ret.}}$

Recall $= \frac{\text{Rel. ret}}{\text{Total rel.}}$

F1 $= \frac{2P.R}{P+R}$

(c) Precision, recall & F1 when doc. is considered relevant ~~only~~ when any 1 or both think it's relevant.

* Precision $= \frac{4}{5} = 0.8$

* Recall $= \frac{4}{6} = 0.666$

* F1 $= \frac{0.8(0.666)2}{0.666+0.8} = 0.72729$

Problem #03

Rocchio's Algorithm formula:

$$\vec{q}_m = \alpha \vec{q}_0 + \frac{\beta}{|D_r|} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \frac{\gamma}{|D_{nr}|} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j$$

$\alpha = 0.1$, $\beta = 0.2$, $\gamma = 0.4.$

Date:

Sum of non-relevant documents $\left(\sum_{\vec{d_j} \in D_{nr}} \vec{d_j}\right) \to d_3$

$$\sum_{\vec{d_j} \in D_{nr}} \vec{d_j} = \langle 0.11, 0.21, 0.11, 0.11, 0.12, 0.12, 0.01 \rangle = X$$

Sum of relevant documents $\left(\sum_{\vec{d_j} \in D_r} \vec{d_j}\right) \to d_1, d_2, d_4.$

$$\sum_{\vec{d_j} \in D_r} \vec{d_j} = \langle 0.61, 0.71, 0.31, 0.21, 0.55, 0.81, 0.41 \rangle = \boxed{Y}$$

$$\vec{q_m} = \alpha \vec{q_0} + X \frac{\beta}{|D_r|} + \gamma \frac{Y}{|D_{nr}|} \quad \text{——} (1)$$

∴

(a) $\vec{q_m} = 0.1 (\langle 0.01, 0.22, 0.11, 0.01, 0.01, 0.22, 0.1 \rangle) +$
$\frac{0.2}{3} (\langle 0.61, 0.71, 0.31, 0.21, 0.55, 0.81, 0.41 \rangle) -$
$\frac{0.4}{1} (\langle 0.11, 0.21, 0.11, 0.11, 0.12, 0.12, 0.01 \rangle)$

$$\vec{q_m} = \langle -3\times10^{-3}, 0.015, -0.013, -0.029, -0.0103, 0.02, 0.0334 \rangle$$

after ignoring the -ve weights we'll get:

$$\vec{q_m} = \langle 0, 0, 0, 0, 0, 0.02, 0.334 \rangle$$

(b) $\vec{q_m} = 0.1 (\langle 0.01, 0.022, 0.11, 0.01, 0.01, 0.22, 0.1 \rangle)$ → for eliminating D_{nr}, we set $\gamma = 0$.
0.067 $+ \frac{0.2}{3} (\langle 0.61, 0.71, 0.31, 0.21, 0.55, 0.81, 0.41 \rangle)$
$= \langle 0.04, 0.069, 0.031, 0.015, 0.0377, 0.072, 0.0374 \rangle$

in this scenario, we can also consider $\alpha = 0.1, \beta = 0.8$ and $\gamma = 0$, then
Solution will be: $\langle 0.05, 0.069, 0.317, 0.15, 0.04, 0.274, 0.127 \rangle$

(c) $\vec{q_m}$ will be same as $\vec{q_0}$ when in eq-(1): $\alpha = 1$
and, $\frac{\beta}{|D_r|} X = \frac{\gamma}{|D_{nr}|} Y$

No, q_0 will be close to centroid of relevant docs if β is very small as compared to γ.