

1) Sia q una query che ha 6 documenti rilevanti nella collezione. Supponiamo che un algoritmo di ritrovamento riporti il seguente ranking R_q (R indica che il documento è rilevante; N indica che il documento è non rilevante; il risultato più a sinistra è il top della lista):

R_q : RRRNNNNRRR

a) Fornire la descrizione sintetica delle metriche: Precision, Recall, Average Precision

(PUNTI 3)

b) Calcolare Precision, Recall ed Average Precision per la query q

(PUNTI 4)

c) Riportare la curva di precision-recall per la query q , usando gli 11 livelli standard di recall

(PUNTI 3)

$$P = \frac{\# \text{ doc rilev. ri-r. veri}}{\# \text{ doc ri-r.}} = \frac{5}{10} = 0.5$$

$$R = \frac{\# \text{ doc ril. ri-r. veri}}{\# \text{ doc ri-r. veri}} = \frac{5}{6}$$

$$AP = \frac{1}{m} \sum_{k=1}^m \text{Precision}(p=k) \quad m = \# \text{ doc ril.}$$

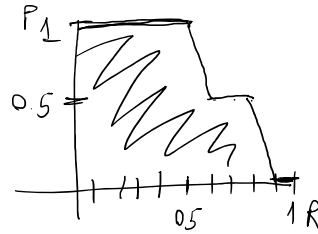
$$AP = \frac{1 + 2/2 + 3/3 + 4/4 + 5/5 + 0}{6} = \frac{4}{6} = \frac{2}{3}$$

Applico procedure di interpolazione per ottenere la Precisione agli 11 punti standard di Recall

$$P(r_j) = \max_{r_i \geq r_j} P(r_i)$$

$P_{0.5}$	P	R
1	1	$1/6 = 0.166$
2	1	$2/6 = 0.333$
3	1	$3/6 = 0.5$
4		
5		
6		
7		
8	$4/8 = 0.5$	$4/6 = 0.666$
9		
10	$5/10 = 0.5$	$5/6 = 0.833$

P	R
1	0
1	0.1
1	0.2
1	0.3
1	0.333
1	0.4
1	0.5
0.5	0.6
0.5	0.666
0.5	0.7
0.5	0.8
0.5	0.833
0	0.9
0	1



1) Sia q una query che ha 6 documenti rilevanti nella collezione. Supponiamo che un algoritmo di ritrovamento applicato a q riporti il seguente ranking R_q : D1 D5 D3 D7 D9 D4
Supponiamo che D1, D3 e D9 siano documenti rilevanti per q

a) Calcolare l'*Average Precision* per la query q , fornendo anche una descrizione della metrica (PUNTI 3)

b) Riportare la curva di precision-recall per la query q , usando gli 11 livelli standard di recall (PUNTI 3)

c) Supponendo di avere dei giudizi di rilevanza non binari, e assumendo che D9 abbia un grado di rilevanza pari a 3, mentre D1 e D3 abbiano un grado di rilevanza pari a 1, calcolare il valore del *DCG* (Discounted Cumulative Gain) per q , fornendo anche una breve descrizione della metrica.

RIL

Document	Relevance	Rank
D1	1	1
D5	0	2
D3	1	3
D7	0	4
D9	3	5
D4	0	6

$AP = \frac{1 + 2/3 + 3/5}{6}$

$P(r_j) = \max P(r)$

$Rec \approx r_j$

$DCG[i] = \begin{cases} G[i] & i=1 \\ \frac{G[i]}{\log_2 i} + DCG[i-1] & i>1 \end{cases}$

$G = (1, 0, 1, 0, 3, 0)$

$DCG = (1, \frac{0+1}{\log_2 2}, \frac{1+1}{\log_2 3}, \frac{0+1+1}{\log_2 4}, \frac{3+1+1}{\log_2 5}, \frac{3+1+1}{\log_2 6})$

Curva Precision-Recall:

P	R
1	0
1	1/6 = 0.166
1	2/6 = 0.333
0.666	3/6 = 0.5
0.666	4/6 = 0.666
0.666	5/6 = 0.833
0.666	6/6 = 1

$$G = (1, 0, 1, 0, 3, 0)$$

$$DCG = (1, \frac{0+1}{\log_2 2}, \frac{1+1}{\log_2 3}, \frac{0+1+1}{\log_2 4}, \frac{3+1+1}{\log_2 5}, \frac{3+1+1}{\log_2 6})$$

CALCOLARE $nDCG$

Prendi il vettore ideale del ranking

$$IG = (3, 1, 1, 0, 0, 0)$$

$$IDCG = (3, \frac{1}{\log_2 2} + 3 = 4, \frac{1}{\log_2 3} + 4, \frac{1}{\log_2 4} + 4, \frac{1}{\log_2 5} + 4, \frac{1}{\log_2 6} + 4)$$

$$nDCG = \frac{DCG}{IDCG} = \left(\frac{1}{3}, \frac{1}{4}, \frac{\frac{1+1}{\log_2 3}}{\frac{1}{\log_2 3} + 4}, \dots \right)$$

R_{q_1}

#doc $|r_i| = 5$

R_{q_2}

$$AP_{q_1} = \frac{1 + 2/3 + 3/4 + 4/6 + 5/8}{5} \approx 0.74$$

$$AP_{q_2} = \frac{1/2 + 2/3 + 3/4 + 4/7 + 5/10}{5} \approx 0.596$$

x
o
x
x
o
x
o
o
o

o
x
x
x
o
o
x
o
x
o
x

q_1

P

R

1

$$1/5 = 0.2$$

$$2/3 = 0.667$$

$$2/5 = 0.4$$

$$3/4 = 0.75$$

$$3/5 = 0.6$$

$$4/6 = 0.667$$

$$4/5 = 0.8$$

$$5/8 = 0.62$$

$$5/5 = 1$$

$$P(r_j) = \max_{\forall r_i \geq r_j} P(r_i)$$

P

R

1

o

1

o1

1

o2

0.75

o3

0.75

o4

0.75

o5

0.75

o6

0.667

o7

0.667

o8

0.62

o9

0.62

1

$$R\text{-precision}_{q_1} = \frac{3}{5}$$

$$R\text{-precision}_{q_2} = \frac{3}{5}$$