

# Web Search Engines — Homework 3

Anirudhan J Rajagopalan

N18824115

ajr619

March 28, 2016

## 1 Problem 1

The decision of the two judges can be summarized as:

		Judge 2	
		Yes	No
Judge 1	Yes	2	4
	No	4	2

For the above condition, we get

$$P(A) = \frac{2+2}{12}$$

$$P(\text{nonrelevant}) = \frac{6+6}{12+12} = 0.5$$

$$P(\text{relevant}) = \frac{6+6}{12+12} = 0.5$$

$$\begin{aligned} P(E) &= P(NR)^2 + P(R)^2 \\ &= 0.5^2 + 0.5^2 \\ &= 0.5 \end{aligned}$$

### 1.1 kappa statistic

$$\begin{aligned} \kappa &= \frac{P(A) - P(E)}{1 - P(E)} \\ &= \frac{0.33 - 0.5}{1 - 0.5} \\ &= -0.36 \end{aligned}$$

### 1.2 Relevant only if both judges agree

#Relevant documents retrieved = 1

#retrieved documents = 5

#relevant documents = 4

#### 1.2.1 Precision

$$\begin{aligned} \text{Precision} &= \frac{\text{\#relevantdocumentsretrieved}}{\text{\#retrieveddocuments}} \\ &= \frac{1}{5} \\ &= 0.2 \end{aligned}$$

**1.2.2 Recall**

$$\begin{aligned}
 Recall &= \frac{\#relevantdocumentsretrieved}{\#relevantdocuments} \\
 &= \frac{1}{4} \\
 &= 0.25
 \end{aligned}$$

**1.2.3  $F_1$  — score**

$$\begin{aligned}
 F_1 &= \frac{2PR}{P + R} \\
 &= \frac{2 * 0.2 * 0.25}{0.2 + 0.25} \\
 &= 0.222
 \end{aligned}$$

**1.3 Relevant if either judge thinks relevant**

#Relevant documents retrieved = 5

#retrieved documents = 5

#relevant documents = 10

**1.3.1 Precision**

$$\begin{aligned}
 Precision &= \frac{\#relevantdocumentsretrieved}{\#retrieveddocuments} \\
 &= \frac{5}{5} \\
 &= 1
 \end{aligned}$$

**1.3.2 Recall**

$$\begin{aligned}
 Recall &= \frac{\#relevantdocumentsretrieved}{\#relevantdocuments} \\
 &= \frac{5}{10} \\
 &= 0.5
 \end{aligned}$$

**1.3.3  $F_1$  — score**

$$\begin{aligned}
 F_1 &= \frac{2PR}{P + R} \\
 &= \frac{2 * 1 * 0.5}{1 + 0.5} \\
 &= 0.667
 \end{aligned}$$

**2 Problem 2****2.1 How to evaluate an image search engine**

Evaluating any search engine involves defining what “relevancy” is with respect to the particular audience the search engine is designed to serve. The relevancy of a search engine might be dependent on how the input is provided. In case of image search, the search can be done by text or by providing images.

**2.1.1 Image input**

When a search is based on image input, the returned images should be similar to the input image. This will depend on various descriptors of an image such as brightness, contrast, colors and so on.

Suppose we upload an image of a patch of tiger skin, the first relevant answer can be a sofa which has stripes similar to that of a tiger. This result is acceptable because the sofa’s pattern matches the input image’s pattern.

**2.1.2 Text input**

When an image search is triggered by textual input, the result should be an image which best describes the textual content given by the user. The most important way in which we can check for relevancy is to verify if objects representing each and every search term is present in the result images.

Suppose an user searches for “Monkey swinging from a tree”. The first result should have a monkey, and a tree and possibly the monkey swinging. The second result can have a monkey, a tree and a swing and so on.

Suppose, the user searches for “tiger skin”, the first relevant answer should be an image which shows tiger skin and not a sofa which is covered by a cloth similar to tiger skin.

**2.2 How evaluating image search differs from textual search**

As discussed above, evaluating image search differs from textual search mainly because of the type of inputs that can be provided to the system.

Also the concept of relevancy can get confusing when we deal with images & text, especially when the input is an image. An image search for “August Rush” can return the posters of movie *August rush* which is quite relevant for

people looking for the movie. But it can also return images of a festival which happens in August and attracts huge amount of people. This may or maynot be relevant to the user who tries to search for images using textual content.

### 3 Problem 3

Given a relation between Gold standard and custom clustering as follows:

	A	B	C
X	10	40	20
Y	50	10	20

#### 3.1 Purity of cluster

For the given example

$$\begin{aligned}\Omega &= X, Y \\ \mathbb{C} &= A, B, C\end{aligned}$$

The purity is given by

$$purity(\Omega, \mathbb{C}) = \frac{1}{N} \sum_k \max_j |\omega_k \cap C_j|$$

We interpret  $\omega_k$  as the set of documents in  $\omega_k$  and  $c_j$  as the set of documents in  $c_j$ .

Therefore the purity for the given cluster combination is given by

$$\begin{aligned}purity(\Omega, \mathbb{C}) &= \frac{1}{150} (40 + 50) \\ &= \frac{90}{150} \\ &= 0.6\end{aligned}$$

#### 3.2 Computation for TP, TN, FP, FN, P, R, $F_\beta$

##### 3.2.1 TP — True Positives

Pairs  $\alpha - \beta$  which are in the same cluster in both the gold standard and in the algorithmic clustering.

The True Positives are given by

$$\begin{aligned}TP &= \frac{10 * 9}{2} + \frac{40 * 39}{2} + \frac{20 * 19}{2} + \frac{50 * 49}{2} + \frac{10 * 9}{2} + \frac{20 * 19}{2} \\ &= 20 * 39 + 25 * 49 + 20 * 19 + 10 * 9 \\ &= 2475\end{aligned}$$

**3.2.2 FP — False Positives**

Pairs  $\alpha - \beta$  which are in the same cluster in the algorithmic cluster but different clusters in the gold standard.

The False Positives are given by

$$\begin{aligned} FP &= 10 * 40 + 10 * 20 + 40 * 20 + 50 * 10 + 50 * 20 + 10 * 20 \\ &= 400 + 200 + 800 + 500 + 1000 + 200 \\ &= 3100 \end{aligned}$$

**3.2.3 FN — False Negatives**

Pairs  $\alpha - \beta$  which are in the same cluster in the gold standard, but different algorithmic clusters.

The False Negatives are given by

$$\begin{aligned} FN &= 10 * 50 + 40 * 10 + 20 * 20 \\ &= 500 + 400 + 400 \\ &= 1300 \end{aligned}$$

**3.2.4 TN — True Negatives**

Pairs  $\alpha - \beta$  that are in different clusters in both the gold standard and the algorithmic clustering.

The True Negatives are given by

$$\begin{aligned} TN &= 10 * 10 + 10 * 20 + 40 * 50 + 40 * 20 + 20 * 50 + 20 * 10 \\ &= 100 + 200 + 2000 + 800 + 1000 + 200 \\ &= 4300 \end{aligned}$$

**3.3 Rand index**

$$\begin{aligned} RI &= \frac{TP + TN}{TP + FP + FN + TN} \\ &= \frac{6775}{11175} \\ &= 0.606 \end{aligned}$$

### 3.4 Precision

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \\
 &= \frac{2475}{2475 + 3100} \\
 &= \frac{2475}{5575} \\
 &= 0.4439
 \end{aligned}$$

### 3.5 Recall

$$\begin{aligned}
 Recall &= \frac{TP}{TP + FN} \\
 &= \frac{2475}{2475 + 1300} \\
 &= \frac{2475}{3775} \\
 &= 0.6556
 \end{aligned}$$

### 3.6 F score when $\beta=1$

$$\begin{aligned}
 F_\beta &= \frac{(\beta^2 + 1) * P * R}{\beta^2 P + R} \\
 &= \frac{2PR}{P + R} \\
 &= \frac{2 * 0.4439 * 0.6556}{0.4439 + 0.6556} \\
 &= \frac{0.5820}{1.0995} \\
 &= 0.5293
 \end{aligned}$$