Due: 2018/05/04

Homework 8

Problem 1. (20 points) An IR system returns 4 relevant documents, and 10 nonrelevant documents. There are a total of 20 relevant documents in the collection. What is the precision of the system on this search, and what is its recall?

Solution:
Precision = $\frac{8}{18} = 0.44$
Recall = $\frac{8}{20} = 0.4$

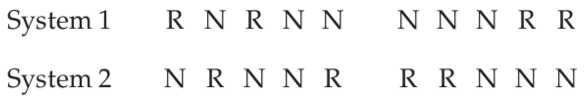
Problem 2. (20 points) The Dice coefficient of two sets is a measure of their intersection scaled by their size (giving a value in the range 0 to 1):

$$Dice(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

Show that the balanced F-measure (F1) is equal to the Dice coefficient of the retrieved and relevant document sets.

Solution:
$F_1 = \frac{2PR}{P+R}$
where $P = \frac{tp}{tp+fp}$
$R = \frac{tp}{tp+fn} \implies F_1 = \frac{2*tp}{2tp+fp+fn}$
$Dice(X,Y) = \left|\frac{X \cap Y}{|X|+|Y|}\right|$
where X = set of retrieved documents
Y = set of relevant documents
$|X| = tp + fp$
$|Y| = tp + fn \implies Dice(X,Y) = \frac{tp}{2tp+fp+fn}$

Problem 3. (30 points) Consider an information need for which there are 4 relevant documents in the collection. Contrast two systems run on this collection. Their top 10 results are judged for relevance as follows (the leftmost item is the top ranked search result):



System 1	R	N	R	N	N	N	N	N	R	R
System 2	N	R	N	N	R	R	R	N	N	N

a. What is the MAP of each system? Which has a higher MAP?
b. Does this result intuitively make sense? What does it say about what is important in getting a good MAP score?
c. What is the R-precision of each system? (Does it rank the systems the same as MAP?)

Solution:
System 1:
(A) MAP= $(\frac{1}{4}) * (1 + (\frac{2}{3}) + (\frac{3}{9}) + (\frac{4}{10})) = 0.6$
(B) The text says that MAP provides a single figure measure of quality across recall levels. It rewards for relevant documents in the beginning of the list, and that can be seen as beneficial. For a good MAP score, it is essential to more relevant documents in the first few retrieves.
(C) R-Precision $= \frac{1}{2}$
System 2:
(A) $(\frac{1}{4}) * (\frac{1}{2} + \frac{2}{5} + \frac{3}{6} + \frac{4}{7}) = 0.493$
(B) The text says that MAP provides a single figure measure of quality across recall levels. It rewards for relevant documents in the beginning of the list, and that can be seen as beneficial. For a good MAP score, it is essential to more relevant documents in the first few retrieves.
(C) R-Precision $=\frac{1}{4}$
Conclusion
(A) System 1 has higher average precision then System 2.
(B) For a good MAP score, it is essential to more relevant documents in the first few retrieves.
(C) R-Precisions of both the systems states that the ranks of the system are as same as MAP.

Problem 4. (30 points) Below is a table showing how two human judges rated the relevance of a set of 12 documents to a particular information need (0 = nonrelevant, 1 = relevant). Let us assume that youve written an IR system that for this query returns the set of documents [4, 5, 6, 7, 8].

docID	Judge 1	Judge 2
1	0	0
2	0	0
3	1	1
4	1	1
5	1	0
6	1	0
7	1	0
8	1	0
9	0	1
10	0	1
11	0	1
12	0	1

a. Calculate the kappa measure between the two judges.
b. Calculate precision, recall, and F1 of your system if a document is considered relevant only if the two judges agree.
c. Calculate precision, recall, and F1 of your system if a document is considered relevant if either judge thinks it is relevant.

Solution:
(A). P(Agree) = $\frac{4}{12}$
P(nonrelevant) = $\frac{12}{24}$
p(relevant) = $\frac{12}{24}$
P(E) = $0.52 + 0.52 = 0.5$
K = $\frac{(0.33-0.5)}{0.5} = -0.34$

(B). P =$\frac{1}{5}$
R = $\frac{1}{2}$
F = $2\ \ \frac{1}{5}\ \frac{\frac{1}{2}}{(\frac{1}{5}+\frac{1}{2})} = \frac{2}{7}$

(C). P = 1
R = $\frac{1}{2}$

F = $\frac{2}{3}$