Exercise 7.3
If we were to only have one-term queries, explain why the use of global champion lists with $r = K$ suffices for identifying the K highest scoring documents. What is a simple modification to this idea if we were to only have s-term queries for any fixed integers $s > 1$?

Solution

1. We take the union of the champion lists for each of the terms comprising the query, and restrict cosine computation to only the documents in the union set. If the query contains only one term, we just take the list with r = K, because there is no need to compute the union.
2. For each term, identify the $\left\lceil \frac{K}{s} \right\rceil$ highest scoring documents.

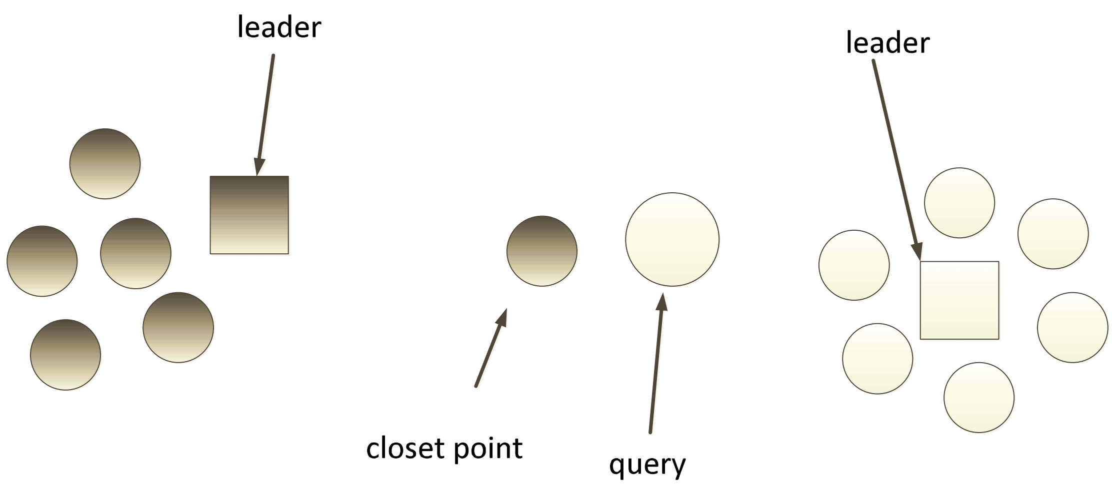
Exercise 7.5
Consider again the data of Exercise 6.23 with `nnn.atc`for the query-dependent scoring. Suppose that we were given static quality scores of 1 for Doc1 and 2 for Doc2.Determine under Equation (7.2) what ranges of static quality score for Doc3 result in it being the first, second or third result for the query best car insurance.

Solution
Suppose the static quality score for Doc3 is g(doc3).
According to Exercise 6.23 and Equation 7.2, score(doc1, q) = 0.7627+ 1 = 1.7627, score(doc2, q) = 0.6839 + 2 = 2.6839, score(doc3, q) = 0.9211 + g(doc3).
For Doc3 result in being:
(1) the first: 0.9211 + g(doc3) > 2.6839, we get g(doc3) > 1.7628
(2) the second: 1.7627< 0.9211 + g(doc3) < 2.6839, we get 0.8416 < g(doc3) < 1.7628
(3) the third: 0.9211+ g(doc3) <1.7627, we get 0 <=g(doc3) < 0.8416.

Exercise 7.8
The nearest-neighbor problem in the plane is the following: given a set of N data points on the plane, we preprocess them into some data structure such that, given a query point Q, we seek the point in N that is closest to Q in Euclidean distance. Clearly cluster pruning can be used as an approach to the nearest-neighbor problem in the plane, if we wished to avoid computing the distance from Q to every one of the query points. Devise a simple example on the plane so that with two leaders, the

answer returned by cluster pruning is incorrect (it is not the data point closest to *Q*).

Solution



As is shown in the above picture, the right leader is closer to the query point than the left leader, but the closest point belongs to the left group.

Exercise 8.1
An IR system returns 8 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 = 8/18 = 0.44
Recall = 8/20 = 0.4

Exercise 8.3
Derive the equivalence between the two formulas for F measure shown in Equation (8.5), given that $\alpha = 1/(\beta^2 + 1)$.

$$F = \frac{1}{\alpha\frac{1}{P} + (1-\alpha)\frac{1}{R}} = \frac{1}{\frac{1}{\beta^2+1}\frac{1}{P} + \frac{\beta^2}{\beta^2+1}\frac{1}{R}} = \frac{(\beta^2+1)PR}{\beta^2 P + R}$$

Exercise 8.9
The following list of Rs and Ns represents relevant (R) and nonrelevant (N) returned documents in a ranked list of 20 documents retrieved in response to a query from a collection of 10,000 documents. The top of the ranked list (the document the system thinks is most likely to be relevant) is on the left of the list. This list shows 6 relevant documents. Assume that there are 8 relevant documents in total in the collection.
R R N NNNNN R N R N NN R N NNN R

a. What is the precision of the system on the top 20?
b. What is the F_1 on the top 20?
c. What is the uninterpolated precision of the system at 25% recall?
d. What is the interpolated precision at 33% recall?
e. Assume that these 20 documents are the complete result set of the system. What is the MAP for the query?
Assume, now, instead, that the system returned the entire 10,000 documents in a ranked list, and these are the first 20 results returned.
f. What is the largest possible MAP that this system could have?
g. What is the smallest possible MAP that this system could have?
h. In a set of experiments, only the top 20 results are evaluated by hand. The result in (e) is used to approximate the range (f)–(g). For this example, how large (in absolute terms) can the error for the MAP be by calculating (e) instead of (f) and (g) for this query?

Solution

a. Precision = 6/20 = 0.3
b. Recall = 6/8 = 0.75

$$F_1 = \frac{2PR}{P+R} = \frac{3}{7} = 0.43$$

c. 8*0.25 = 2, the uninterpolated precision could be 1, 2/3, 2/4, 2/5, 2/6, 2/7, 1/4
d. Because the highest precision found for any recall level larger than 33% is 4/11 = 0.364, hence the interpolated precision at 33% recall is 4/11 = 0.364.

e. MAP = 1/6 * (1 + 1 + 3/9 + 4/11 + 5/15 + 6/20) = 0.555

f. $\text{MAP}_{\text{largest}} = \frac{1}{8} * \left(1 + 1 + \frac{3}{9} + \frac{4}{11} + \frac{5}{15} + \frac{6}{20} + \frac{7}{21} + \frac{8}{22}\right) = 0.503$

g. $\text{MAP}_{\text{smallest}} = \frac{1}{8} * \left(1 + 1 + \frac{3}{9} + \frac{4}{11} + \frac{5}{15} + \frac{6}{20} + \frac{7}{9999} + \frac{8}{10000}\right) = 0.417$

h. 0.555-0.417 = 0.138, 0.555-0.503 = 0.052
the error is in [0.052, 0.138]

Exercise 8.10

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 you've 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(A) = 4/12 = 1/3

P(nonrelevant) = (6+6)/(12+12) = 0.5, P(relevant) = (6+6)/(12+12) = 0.5

P(E)=0.5*0.5 + 0.5*0.5 = 0.5

Kappa = (P(A)-P(E))/(1-P(E)) = -1/3

(b)

Precision = 1/5 = 0.2

Recall = 1/2 = 0.5

F1 = 2*0.2*0.5/(0.2+0.5) = 2/7 = 0.286

(c)

Precision = 5/5 = 1

Recall = 5/10 = 0.5

F1 = 2*1*0.5/(1+0.5) = 2/3 = 0.667