Homework 1: Evaluation Metrics

Student ID: 16340024 Student Name:陈铭涛

Information Retrieval Course
Sun Yat-sen University

Exercise 1: Rank-based Evaluation Metrics

Assume the following ranking for a given query (only results 1-10 are shown). The column 'rank' gives the rank of the document. The column 'docID' gives the document ID associated with the document at that rank. The column 'graded relevance' gives the relevance grade associated with the document (4 = perfect, 3 = excellent, 2 = good, 1 = fair, and 0 = bad). The column 'binary relevance' provides two values of relevance (1 = relevant and 0 = non-relevant). The assumption is that anything with a relevance grade of 'fair' or better is relevant and that anything with a relevance grade of 'bad' is non-relevant.

Also, assume that this query has only 7 documents with a relevance grade of fair or better. All happen to be ranked within the top 10 in this given ranking.

Answer the questions below. P@K, R@K, and average precision (AP) assume binary relevance. For those metrics, use the 'binary relevance' column. DCG and NDCG assume graded relevance. For those metrics, use the 'graded relevance' column.

rank	docID	graded relevance	binary relevance
1	51	4	1
2	501	1	1
3	21	0	0
4	75	3	1
5	321	4	1
6	38	1	1
7	521	0	0
8	412	1	1
9	331	0	0
10	101	2	1

(a) Compute P@5 and P@10.

$$P@5 = \frac{4}{5} = 0.8$$
$$P@10 = \frac{7}{10} = 0.7$$

(b) Compute R@5 and R@10.

$$R@5 = \frac{4}{7} = 0.57$$
$$R@10 = \frac{7}{7} = 1$$

(b) Provide an example ranking for this query that maximizes P@5.

rank	docID
1	51
2	501
3	75
4	321
5	38
6	412
7	101
8	21
9	521

10	331

$$P@5 = \frac{5}{5} = 1$$

(d) Provide an example ranking for this query that maximizes P@10.

rank	docID
1	51
2	501
3	75
4	321
5	38
6	412
7	101
8	21
9	521
10	331

$$P@10 = \frac{7}{10} = 0.7$$

(e) Provide an example ranking for this query that maximizes R@5.

rank	docID
1	51
2	501
3	75
4	321
5	38
6	412
7	101
8	21
9	521
10	331

$$R@5 = \frac{5}{7} = 0.71$$

(f) Provide an example ranking for this query that maximizes R@10.

rank	docID
1	51
2	501
3	75
4	321
5	38
6	412
7	101
8	21
9	521
10	331

$$R@10 = \frac{7}{7} = 1$$

(g) You have reason to believe that the users of this system will want to examine every relevant document for a given query. In other words, you have recent to believe that users want perfect recall. You want to evaluate based on P@K. Is there a query-specific method for setting the value of K that would be particularly appropriate in this scenario? What is it? (**Hint**: there is an evaluation metric called R-Precision, which we did not talk about in the lectures. Your answer should be related to R-Precision. Wikipedia might help.)

R-Precision 即 P@X，X 是所有 relevant 的文档数量，对于当前 query，R-Precision 为：

$$R-Precision=\frac{r}{X}=\frac{5}{7}=0.71$$

R-Precision 同时与 R@X 相等，因此为获得完美 recall，$K=\frac{X}{R-Precision}$为合适的选择。

(h) Compute average precision (AP). What are the difference between AP and MAP (Mean Average precision)?

$$AP=\frac{1}{7}\left(1+1+\frac{3}{4}+\frac{4}{5}+\frac{5}{6}+\frac{6}{8}+\frac{7}{10}\right)=0.833$$

AP 指的是对于每个 relevant 的文档位置的平均 Precision，MAP 则是对于多组 queries/rankings 的平均 Precision

(i) Provide an example ranking for this query that maximizes average precision (AP).

rank	docID
1	51
2	501
3	75
4	321
5	38
6	412
7	101
8	21
9	521
10	331

$$AP=\frac{1}{7}(1+1+1+1+1+1+1)=1$$

(j) Compute $NDCG_5$ (i.e., the discounted cumulative gain at rank 5).

$$DCG_5=4+\frac{1}{\log(2)}+\frac{3}{\log(4)}+\frac{4}{\log(5)}=8.223$$

$$IDCG_5=4+\frac{4}{\log(2)}+\frac{3}{\log(3)}+\frac{2}{\log(4)}+\frac{1}{\log(5)}=11.323$$

$$NDCG_5=\frac{DCG_5}{IDCG_5}=0.7262$$

(k) $NDCG_5$ is given by

$$NDCG_5=\frac{DCG_5}{IDCG_5}$$

where *IDCG*$_5$ is the *DCG*$_5$ associated with the *ideal* top-5 ranking associated with this query. Computing *NDCG*$_5$ requires three steps.

(i) What is the *ideal* top-5 ranking associated with this query (notice that the query has 2 *perfect* documents, 1 *excellent* documents, 1 *good* document, 3 *fair* documents, and the rest of the documents are *bad*)?

rank	docID	Graded Relevance
1	51	4
2	321	4
3	75	3
4	101	2
5	501	1

(ii) $IDCG_5$ is the DCG_5 associated with the *ideal* ranking. Compute $IDCG_5$. (**Hint:** compute DCG_5 for your ranking proposed in part (i).)

$$IDCG_5 = 4 + \frac{4}{\log(2)} + \frac{3}{\log(3)} + \frac{2}{\log(4)} + \frac{1}{\log(5)} = 11.323$$

(iii) Compute $NDCG_5$ using the formula above.

$$NDCG_5 = \frac{DCG_5}{IDCG_5} = 0.7262$$

(l) Are there other evaluation metrics to be used to evaluate the performance of the rankings in the table? What are the evaluation scores obtained by these metrics?

Reciprocal rank: 倒数排名是第一个正确答案的倒数积 $RR = \frac{1}{rank} = 1$

F1 Score: $F_1 = \frac{2(P \times R)}{P+R} = 0.824$

Exercise 2: Precision-Recall Curves

A Precision-Recall (PR) curve expresses precision as a function of recall. Usually, a PR-curve is computed for each query in the evaluation set and then averaged. For simplicity, the goal in this question is to draw a PR-curve for a *single* query. Draw the PR-curve associated with the ranking above (same query, same results). **Hint:** Your PR curve should always go down with increasing levels of recall.