

# Homework 2: Evaluation Metrics

Student ID	Student Name
18342075	米家龙

Lectured by: Shangsong Liang

Information Retrieval Course  
Sun Yat-sen University

## Homework 2: Evaluation Metrics

Exercise 1: Rank-based Evaluation Metrics, MAP@K, MRR@K

- (a) AP@5 AP@10 RR@5 RR@10
- (b) MAP@5 MAP@10 MRR@5 MRR@10

Exercise 2: Rank-based Evaluation Metrics, Precision@K, Recall@K, NDCG@K

- (a) P@5 P@10
- (b) R@5 R@10
- (c) maximize P@5
- (d) maximize P@10
- (e) maximize R@5
- (f) maximize R@10
- (g) R-Precision
- (h) AP; difference between AP and MAP
- (i) maximize AP
- (j)  $DCG_5$
- (k)  $NDCG_5$

Exercise 3: Precision-Recall Curves

Exercise 4: Other Evaluation Metrics

- AUC (Area under ROC curve)
- Kendall tau distance
- Spearman's  $\rho$

## Exercise 1: Rank-based Evaluation Metrics, MAP@K, MRR@K

(a) AP@5 AP@10 RR@5 RR@10

query	AP@5	AP@10	RR@5	RR@10
1	0.8333	0.6476	1	1
2	1	0.6429	1	1
3	0	0.2508	0	0.1667

(b) MAP@5 MAP@10 MRR@5 MRR@10

MAP@5	MAP@10	MRR@5	MRR@10
0.6111	0.5138	0.6667	0.7222

## Exercise 2: Rank-based Evaluation Metrics, Precision@K, Recall@K, NDCG@K

---

(a) P@5 P@10

P@5	P@10
0.8000	0.7000

(b) R@5 R@10

由于计算召回率需要数据库的其他数据，但是这里只给了部分搜索结果，因此假设数据库总量就是上述结果

R@5	R@10
4/7	1

(c) maximize P@5

rank	docID	binary relevance
1	51	1
2	501	1
4	75	1
5	321	1
6	38	1

(d) maximize P@10

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

### (e) maximize R@5

rank	docID	binary relevance
1	51	1
2	501	1
4	75	1
5	321	1
6	38	1

$$R@5_{max} = 0.71$$

### (f) maximize R@10

和 (d) 中一样的排序

### (g) R-Precision

R-Precision 是序列前 R 个位置的准确率；为了保证用户的体验，我们需要尽量让 R-Precision 率大

### (h) AP; difference between AP and MAP

$$AP = \frac{(1 + \frac{2}{2} + \frac{3}{4} + \frac{4}{5} + \frac{5}{7} + \frac{6}{8} + \frac{7}{9})}{7} = 0.8333$$

区别：AP 是对一个查询的平均，MAP 则是针对多个查询的 AP 取平均值

(i) maximize AP

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

(j)  $DCG_5$

$DCG_p$  公式采用的是  $\sum_{i=1}^p \frac{rel_i}{\log_2(i+1)}$

$$DCG_5 = \sum_{i=1}^5 \frac{rel_i}{\log_2(i+1)} = 4 + 0.6309 + 0 + 1.2920 + 1.5474 = 7.4703$$

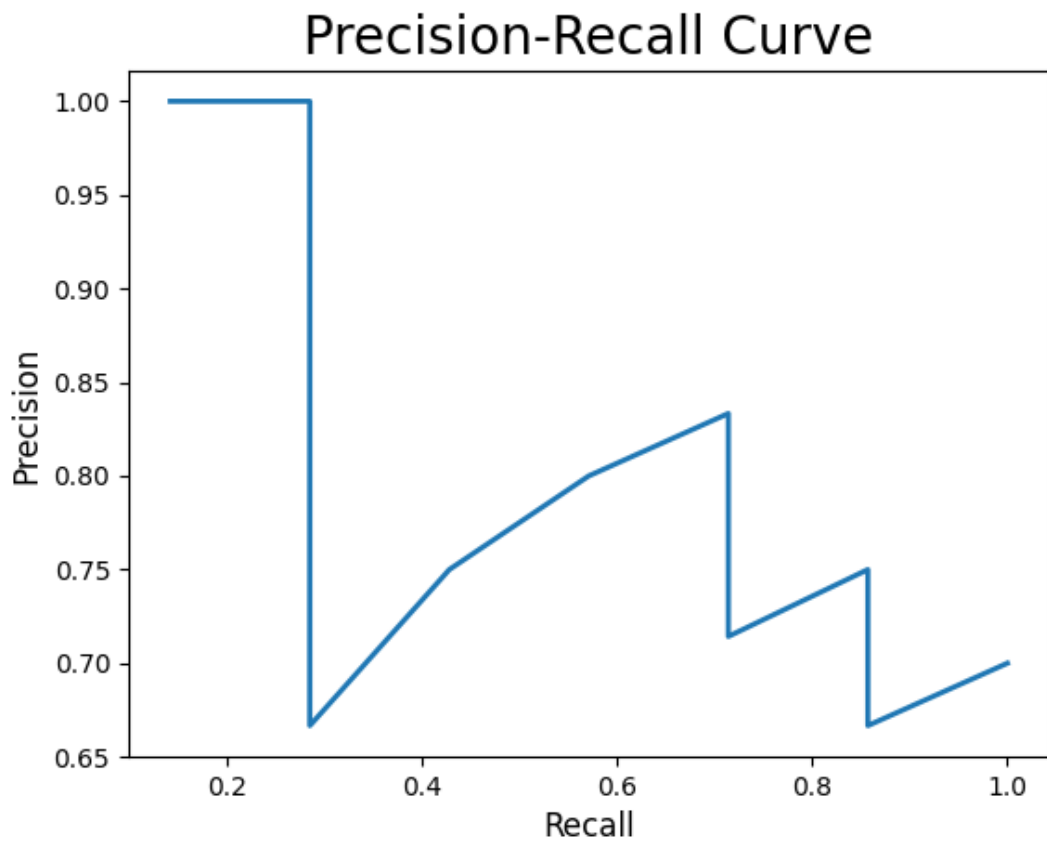
(k)  $NDCG_5$

$$NDCG_5 = \frac{DCG_5}{IDCG_5} = \frac{4 + 0.6309 + 0 + 1.2920 + 1.5474}{4 + 2.523 + 1.5 + 0.8614 + 0.3868} = 0.8056$$

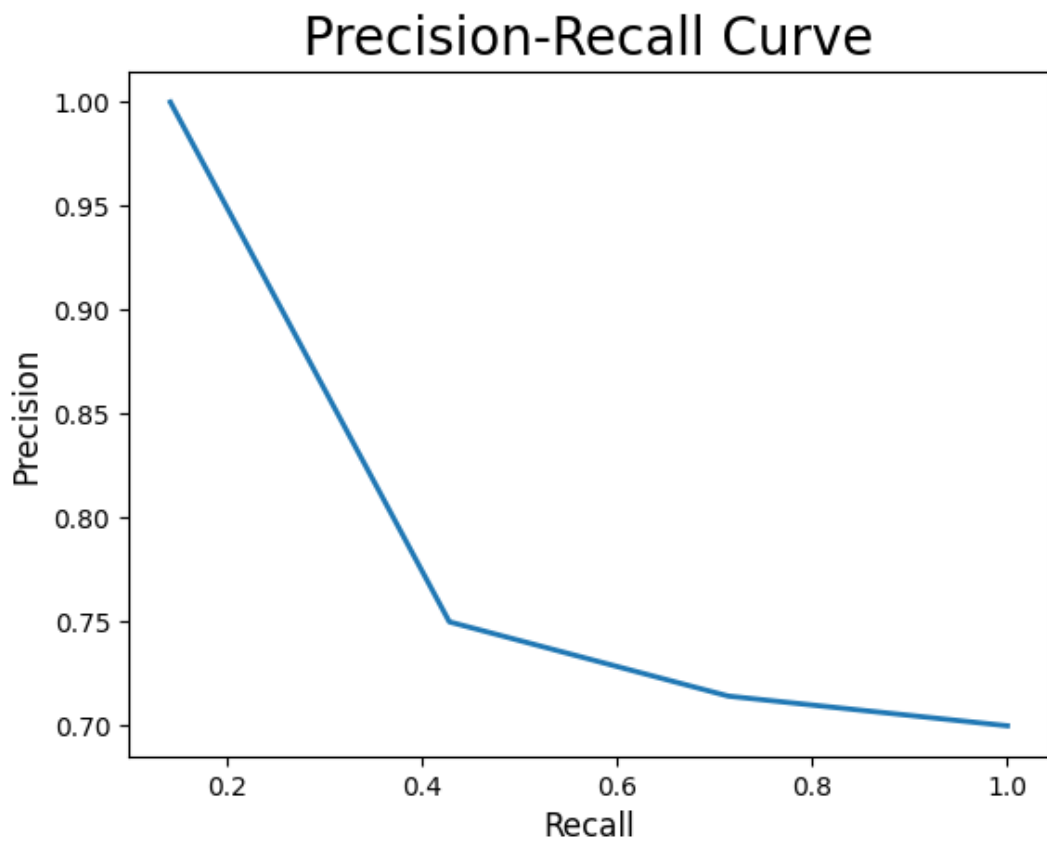
## Exercise 3: Precision-Recall Curves

---

使用全部的数据，发现无法满足要求，如图：



选择使用第 1、4、7、10 次的查询数据，做出如下图：



## Exercise 4: Other Evaluation Metrics

---

## AUC (Area under ROC curve)

AUC的物理意义为任取一对例和负例，正例得分大于负例得分的几率，AUC越大，代表方法效果越好。  
(AUC的值通常介于0.5~1)

## Kendall tau distance

比较两个排序之间，评价存在分歧的对的数量。

$$K(\tau_1, \tau_2) = |\{(i, j) : i < j, (\tau_1(i) < \tau_1(j) \wedge \tau_2(i) > \tau_2(j)) \vee (\tau_1(i) > \tau_1(j) \wedge \tau_2(i) < \tau_2(j))\}|$$

其中  $\tau_1(i)$  和  $\tau_2(i)$  分别表示元素  $i$  在两个排序中的位置

如果两个排序完全一样，那么 Kendall tau distance 为0；如果完全相反，那么为  $n(n-1)/2$ ；通常该距离都会除以  $n(n-1)/2$  来进行归一化

## Spearman's $\rho$

基本思想类似Kendall tau distance：比较两个排序（通常一个是理想排序）的（排序值的）皮尔逊相关系数

$$\frac{\sum_{(i,j) \in \Omega^{test}} (S_{ij}^* - \bar{s}^*)(y_{ij}^* - \bar{y}^*)}{\sqrt{\sum_{(i,j) \in \Omega^{test}} (S_{ij}^* - \bar{s}^*)^2} \sqrt{\sum_{(i,j) \in \Omega^{test}} (y_{ij}^* - \bar{y}^*)^2}}$$

其中  $s_{ij}^*$  表示你模型预测中，物品  $j$  在用户  $i$  的推荐列表上的排序位置； $y_{ij}^*$  表示按实际用户  $i$  对物品的评分来排序时物品  $j$  在  $i$  的推荐列表上的排序位置； $\bar{s}^*$  是  $s_{ij}^*$  的平均值； $\bar{y}^*$  是  $y_{ij}^*$  的平均值