

# 搜索评估指标

## MAP(Mean Average Precision, 平均精确率)

AP即平均精确率是计算返回的结果列表中，各个位置及其之前的文档的精确率的平均值，计算公式写作：

$$AP(q) = \frac{\sum_{i=1}^n Precision_i}{n}$$

其中 $Precision_i$ 是指在前*i*篇文档中的精确率， $n$ 是指Query  $q$ 返回的检索结果条数。

MAP的计算公式为：

$$MAP = \frac{\sum_j^{|Q|} AP_{q_j}}{|Q|}$$

即将所有进行评估的Query求平均值即为MAP值。

## NDCG

NDCG指标是现在搜索领域最常用的评估指标，我们首先来看下DCG的指标计算，公式：

$$DCG_q = \sum_{i=1}^n \frac{2^{(rel)} - 1}{\log(i + 1)}$$

其中 $rel$ 为人工标注的分值，即人工为文档的打分，一般在现实中都是采用5档的方式，即 $rel$ 的取值为[0, 4]，标注的分值越高表明人工认为该结果越相关，我们可以看到该指标我们是可比较不同的排序的模型的效果的，DCG的值越高越好，不过在现实中更常用的是NDCG指标，即对DCG做Normalize，NDCG指标的形式为：

$$NDCG_q = \frac{DCG_q}{IDCG_q}$$

其中的 $IDCG_q$ 即理想的排序所得到的DCG的分值，理想排序就是按照人工打分从大到小排序，即认为该排序是理想排序，并且以模型实际排序计算出DCG除以理想排序的DCG值即得到NDCG。

举例：有3个文档[A, B, C]，人工标注分值为[3, 2, 1]，可以看到该排序即为理想排序，模型排序得到的结果是[B, A, C]，则我们可以计算出模型得到的DCG为：

$\frac{2^2-1}{\log(2)} + \frac{2^3-1}{\log(3)} + \frac{2^1-1}{\log(4)} = 11.42$ ，IDCG为:  $\frac{2^3-1}{\log(2)} + \frac{2^2-1}{\log(3)} + \frac{2^1-1}{\log(4)} = 13.55$ ，则  
NDCG的值为:  $\frac{DCG}{IDCG} = \frac{11.42}{13.55} = 0.84$

## ERR

ERR评估指标是雅虎提出的一个搜索评估方法，其基本的思路是用户看到某条搜索结果不仅和搜索结果的位置有关，还和该搜索结果之前的搜索结果也有关，以此来构建了一个评估方法，其公式比较复杂，下面分几个部分进行：

首先得到Doc让用户满意的概率公式

$$P(d_i) = \frac{2^{(Rel_i)} - 1}{2^{MaxRel}}$$

其中Rel值与NDCG一致也是人工标注的得分，其中MaxRel值即人工标注可以打的最大的值，比如对于评分为[0, 4]的情况，MaxRel的值就是4分， $Rel_i$ 是标识第*i*个文档的标注分值。

在得到单Doc能让用户满意的概率公式后，进一步假设用户只要看到满意结果后就不会继续往下看，则可以得到看到位置*i*用户能够满意的且驻留在位置*i*处Doc的概率是

$$P_i = \prod_{j=1}^i (1 - P(d_j)) P(d_i)$$

即用户对前*i*-1个结果均不满意，在第*i*个位置得到满意结果。

ERR最终的计算公式为：

$$ERR = \sum_{i=1}^n \frac{P_i}{i}$$

我们可以看到， $P(d_i)$ 的值，人工给定的评分越高，则该概率值越大，并且在 $P_i$ 中， $(1 - P(d_i))$ 的值都是介于(0, 1]之间的，所以如果一个评分高的结果越往后排，得到的到该位置让用户满意并驻留的概率就会越小，并且在ERR的计算公式中会除以Doc所在的位置，因此，将一个越好的结果排的越靠后，其在评估中的收益就越小，而将好的结果排的越靠前，则其在评估中的收益就越大。因此，ERR评估指标也是在评估中值越大越好。