

多样性评估指标

多样性评估主要是衡量信息检索系统中结果多样性好坏的指标，下面的几种指标均为Query属于多个类别时的评估指标。但是只要我们适当做些变化用于评估不同召回方式、不同源内容的合并结果的多样性也是行得通的。

NDCG-IA

NDCG-IA的计算公式为：

$$NDCG-IA(Q) = \sum_c P(c|q)NDCG(Q|c)$$

其中c为Query所属的类别，因此，对多样性的评估的就可以看做以不同类别下计算出NDCG与Query属于该类目的概率乘积之和。在计算NDCG(Q|c)时将计算类别不一致的结果看做Bad，也就是将其Rel值置为0。

NDCG的计算公式：

$$NDCG = \frac{\sum_{i=1}^k \frac{2^{Rel(i)} - 1}{\log(i+1)}}{Ideal - DCG}$$

其中的Rel值即为人工标注的分值，一般可以取[0,3]四档，四档分别对应{bad, fair, good, excellent}。从NDCG-IA可以看到当Query只属于一个类别时就会自动退化为NDCG的计算公式，此时也是能够计算出的最大的NDCG值。

MAP-IA

MAP-IA的计算公式：

$$MAP-IA(Q) = \sum_c P(c|q)MAP(Q|c)$$

c的含义与上面NDCG-IA相同，MAP的计算公式为：

$$MAP(Q) = \frac{\sum_q \frac{\sum_{i=1}^k Precision(i)}{k}}{|Q|}$$

上面的两个指标均来源于：Diversifying Search Results.

α -nDCG

α -nDCG是NDCG的变体，其中对新发现的主题的内容进行奖励，多余的主题的内容则进行惩罚。计算公式为：

$$\alpha - nDCG = \frac{\sum_{i=1}^k \frac{G(i)}{\log(i+1)}}{Ideal - DCG}$$

其中 $G(i)$ 的计算公式为：

$$G(i) = \sum_c R(d, c_j)(1 - \alpha)^{r_{c_j, i-1}}$$

其中 $R(d_k, c_j)$ 表示位置 i 处主题为 c_j 的标识，如果为 c_j 则该值取1，否则取0，对公式中的 $r_{c_j, i-1}$ 的计算公式为：

$$r_{c, i-1} = \sum_{l=1}^{i-1} R(d, c_j)$$

表示在当前的前 $i-1$ 个结果中有多少个是主题为 c_j 的，从而我们可以看出某个主题已经被选择的次数越多，则其再次出现是带来的收益就会越小，因为出现次数越多之后折扣就会变得越大。从而通过该指标来衡量检索结果的多样性。

从上面的三个评估指标看， α -nDCG是单纯的进行多样性的评估，而NDCG-IA和MAP-IA不仅仅衡量NDCG，它还考虑了相关性的指标，因此这两个指标在很大程度上反映的是其他主题的结果对概率较大主题结果的补充。