

IR的评价指标-MAP,NDCG和MRR

转载 ullytalk 2016-09-08

MAP(Mean Average Precision):

单个主题的平均准确率是每篇相关文档检索出后的准确率的平均值。主集合的平均准确率(MAP)是每个主题的平均准确率的平均值。MAP 是反映系统在全部相关文档上性能的单值指标。系统检索出来的相关文档越靠前(rank 越高), MAP就可能越高。如果系统没有返回相关文档, 则准确率默认为0。
例如: 假设有两个主题, 主题1有4个相关网页, 主题2有5个相关网页。某系统对于主题1检索出4个相关网页, 其rank分别为1, 2, 4, 7; 对于主题2检索出3个相关网页, 其rank分别为1,3,5。对于主题1, 平均准确率为 $(1/1+2/2+3/4+4/7)/4=0.83$ 。对于主题2, 平均准确率为 $(1/1+2/3+3/5+0+0)/5=0.45$ 。则MAP= $(0.83+0.45)/2=0.64$ 。”

NDCG(Normalized Discounted Cumulative Gain):

计算相对复杂。对于排在结位置n处的NDCG的计算公式如下图所示:

$$N(n) = \underbrace{Z_n}_{\text{Normalization}} \underbrace{\sum_{j=1}^n}_{\text{Cumulating}} \underbrace{(2^{r(j)} - 1)}_{\text{Gain}} \underbrace{/\log(1 + j)}_{\text{Position discount}}$$

在MAP中, 四个文档和query要么相关, 要么不相关, 也就是相关度非0即1。NDCG中改进了下, 相关度分成从0到r的r+1的等级(r可设定)。当取r=5时, 等级设定如下图所示:

Relevance Rating	Value (Gain)
Perfect	31=2 ⁵ -1
Excellent	15=2 ⁴ -1
Good	7=2 ³ -1
Fair	3=2 ² -1
Bad	0=2 ⁰ -1

(应该还有r=1那一级, 原文档有误, 不过这里不影响理解)

例如现在有一个query={abc}, 返回下图左列的Ranked List(URL), 当假设用户的选择与排序结果无关(即每一级都等概率被选中), 则生成的累计增益值如下图最右列所示:

	URL	Gain	Cumulative Gain
#1	http://abc.go.com/	31	31
#2	http://www.abcteach.com/	3	34 = 31 + 3
#3	http://abcnews.go.com/sections/scitech/	15	49 = 31 + 3 + 15
#4	http://www.abc.net.au/	15	64 = 31 + 3 + 15 + 15
#5	http://abcnews.go.com/	15	79 = 31 + 3 + 15 + 15 + 15
#6

考虑到一般情况下用户会优先点选排在前面的搜索结果，所以应该引入一个折算因子(discounting factor): $\log(2)/\log(1+\text{rank})$ 。这时将获得DCG值(Discounted Cumulative Gain)如下所示：

	URL	Gain	Discounted Cumulative Gain
#1	http://abc.go.com/	31	31 = 31x1
#2	http://www.abcteach.com/	3	32.9 = 31 + 3x0.63
#3	http://abcnews.go.com/sections/scitech/	15	40.4 = 32.9 + 15x0.50
#4	http://www.abc.net.au/	15	46.9 = 40.4 + 15x0.43
#5	http://abcnews.go.com/	15	52.7 = 46.9 + 15x0.39
#6

最后，为了使不同等级上的搜索结果的得分值容易比较，需要将DCG值归一化的到NDCG值。操作如下图所示，首先计算理想返回结果List的DCG值：

	URL	Gain	Max DCG
#1	http://abc.go.com/	31	31 = 31x1
#2	http://abcnews.go.com/sections/scitech/	15	40.5 = 31 + 15x0.63
#3	http://www.abc.net.au/	15	48.0 = 40.5 + 15x0.50
#4	http://abcnews.go.com/	15	54.5 = 48.0 + 15x0.43
#5	http://www.abc.org/	15	60.4 = 54.5 + 15x0.39
#6

然后用DCG/MaxDCG就得到NDCG值，如下图所示：

	URL	Gain	DCG	Max DCG	NDCG
#1	http://abc.go.com/	31	31	31	1 = 31/31
#2	http://www.abcteach.com/	3	32.9	40.5	0.81=32.9/40.5
#3	http://abcnews.go.com/sections/scitech/	15	40.4	48.0	0.84=40.4/48.0
#4	http://www.abc.net.au/	15	46.9	54.5	0.86=46.9/54.5
#5	http://abcnews.go.com/	15	52.7	60.4	0.87=52.7/60.4
#6

MRR(Mean Reciprocal Rank):

是把标准答案在被评价系统给出结果中的排序取倒数作为它的准确度，再对所有的问题取平均。相对简单，举个例子：有3个query如下图所示：

Query	Results	Correct response	Rank	Reciprocal rank
cat	catten, cati, cats	cats	3	1/3
torus	torii, tori , toruses	tori	2	1/2
virus	viruses , virii, viri	viruses	1	1

(黑体为返回结果中最匹配的一项)

可计算这个系统的MRR值为： $(1/3 + 1/2 + 1)/3 = 11/18=0.61$ 。