Retrieval System Evaluation Using
Recall-Precision <Problems and solutions>

Concept of Recall:

"Retrieve as many relevant items as possible"

Concept of Precision:

"Retrieve as few non-relevant items as possible"

Recall & Precision are measured after the system determines an ordering on the documents in its collection in response to a user query.

Two problems

(1) When system generates a weak ordering of the documents as the output.
   - Some notions like probability of relevance given retrieved or expected precision have to be introduced.

(2) When a set of queries is involved and we want to evaluate the overall retrieval results based on this given set of queries.
   - Some techniques of interpolation of precision values are needed.

	Rel	NREL	
Ret	a	b	a+b
Not Ret	c	d	
	a+c		

Recall = a / a+c

Precision = a / a+b

Fallout = b / b+d

Often, paired measures are used, e.g. (R, F) pairs or (R, P) pairs.

$$\text{Classification Error} = \frac{b + c}{a + b + c + d}$$

Two types of ordering of the RSL

linear (simple) ordering: Every item in the collection is assigned a distinct RSV by the similarity function used.

weak: More than one item may be present at the same level with identical RSV.

Stopping Criteria: Criterion to stop after retrieving a given # of relevant documents.

(ex) n relevant documents with respect to a given query assuming that the stopping criterion is the retrieval of h relevant documents $1 \leq h \leq n$

➔ $\exists$ n possible recall levels

1/n, 2/n, … h/n, … n-1/n, 1

These are called simple recall levels.

Problem of Weak Ordering

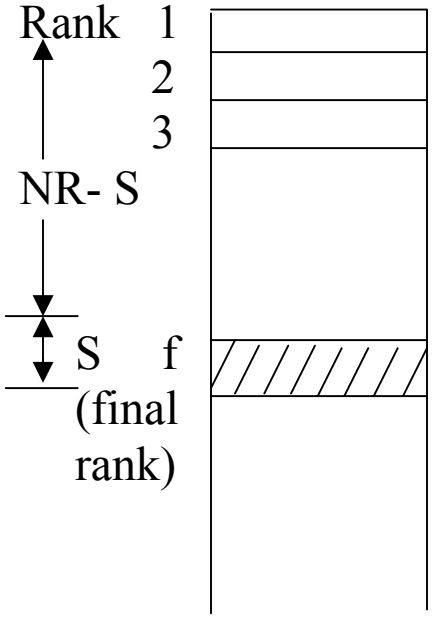
NR: # of relevant documents needed to retrieved for a given query with n relevant documents

$$0 \leq NR \leq n$$

If the ordering is linear, for any recall point NR/n precision is defined as NR/(NR+NNR)

Where NNR= #of non-relevant documents being retrieved along with the NR relevant documents we need.

Weak ordering



Suppose there are r relevant documents and i non-relevant documents at the final rank

It could be imagined that r Relevant docs from r intervals and i nonrelevant docs at the same rank are uniformly distributed among r intervals.

Hence, for every relevant document retrieved i/r non-relevant documents is expected to be retrieved.

$\therefore NNR = j + s*i / r$

➔ j : # of non-relevant docs in ranks above the final rank.

s : # of relevant docs wanted from the final rank.

precision
$= NR /(NR + j + s*i / r) \ldots (1)$

** the problem associated with the above approach is that the validity of the guess concerning the likely distribution of relevant and non-relevant documents at the final rank is questionable.

Problem of Multiple Query

We want to average the precisions of several queues at a recall point.

The simple recall levels (i.e., 1/n, 2/n, …n-1/n, 1) can not be used.

Conventional choice for standardized recall levels is 0, 0.05, 0.1, …, 0.95, 1.

Let x be one of the standardized recall levels
$h/n \leq x \leq (h+1)/n$ and $0 \leq h \leq n$

Then the precision value at point x is assigned the value at the simple recall point (h+1)/n. This method is termed as the ceiling interpolation, since the precision value at the point x*n is same as $\lceil x*n \rceil$

Precision = $\lceil x*n \rceil$ / ($\lceil x*n \rceil$ + j + s*i/r)

This is called Precall with ceiling interploation

** The problem: (1) The precision value by ceiling interpolation does not conform to the general behavior one intuitively expect.
The resulting graph is a step function.

(2)Evaluation results are difficult to interpret.

PRR

Ceiling interpolation

$\lceil x*n \rceil / (\lceil x*n \rceil + j + s*i/r+1)$

preferred interpolation

$x*n / (x*n + j + s*i/r+1)$

PRECALL

Ceiling interpolation

RB-Precision | preferred interpolation |

In case x*n is an integer, ceiling interpolation and preferred interpolation give the same result for precision.

Example:

Δ = (+- -| ± ± _+_)

recall level → 0.5 (s=1)

(2 / 2+2+4/3) = 2 * 3/16 = 3/8

NR/ (NR+NNR)
NNR= j + s*i/r
i = 4
r = 3

recall level → 0.75 (s=2)

3/(3+2+8/3) = 9/23

recall level → 1 (s = 3)

4/ (4+2+4) = 2/5

Recall level → 0.3 (ceiling interpolation)

⌐0.3 * 4¬ /(⌐0.3 * 4¬ +2 +4/3) = 3/8
Note: s=1

Preferred Interpolation

1.2/(1.2+2+0.2*4/3) Note : s = 0.2
=1.2*3 / 10.4
= 3.6 / 10.4
=18 / 52
= 9 / 26