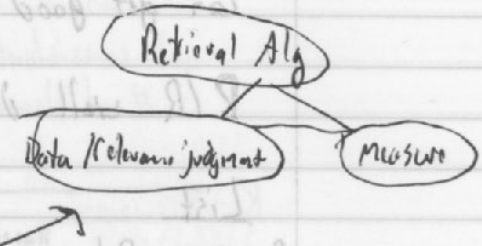
Information Retrieval

5/4/2005 IR models + IR evaluation



Evaluation in document retrieval
- dependent on model for relevance + task

TREC - Text REtrieval Conference

Relevance
- difficult to define
- a relevant document is one judged useful in the context of a query
- w/ real collections, never know full set of relevant documents
- Retrieval model incorporates notion of relevance
- individuals will disagree on individual documents but will generally agree on avg. across data points

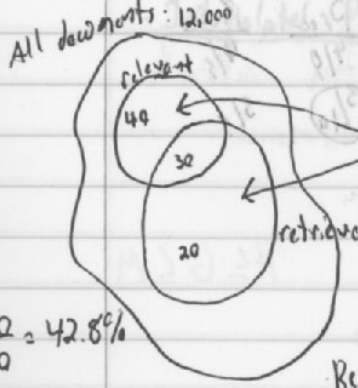
Test Collection
- set of documents
- set of queries
- set of relevance judgements (which docs relevant to each query)

Precision and Recall

|x| - size of x

Precision: proportion of a retrieved set that is relevant

Precision = |relevant ∩ retrieved| ÷ |retrieved| = P(relevant | retrieved)



2 types of errors
- false-positives:
- false negatives.

$Prec = \frac{|retrieved \cap relevant|}{|retrieved|}$

$Prec = \frac{30}{50} = 60\%$

$Recall = \frac{30}{70} = 42.8\%$

Recall: proportion of all relevant documents in the collection included in the retrieved set

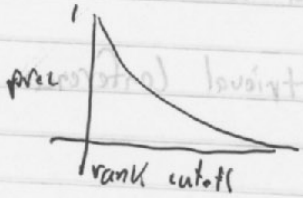
$Recall = \frac{|relevant \cap retrieved|}{|relevant|} = P(retrieved \mid relevant)$

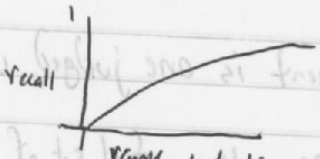
can get good recall at expense of precision: return large (all) set
can get good precision at expense of recall: return 1 relevant doc

R/R well defined for sets

List

Rank	Relevance
1	R
2	N
3	N
4	R
:	R
	:



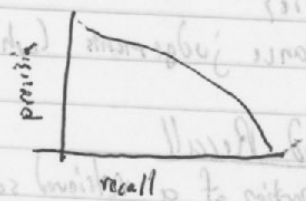


expect precision to drop rank to rise

Average Precision

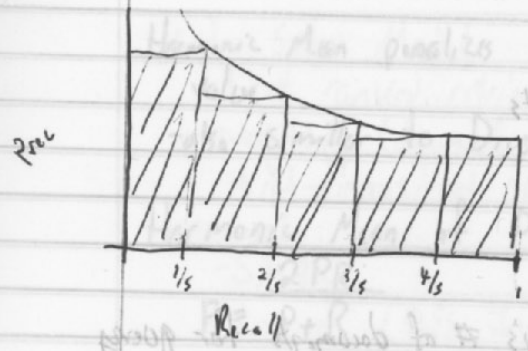
- average precision whenever recall increases (every time find a relevant document)
- highly correlated to what people actually feel
- highly weights early relevant documents in list.

Recall / Precision Graphs

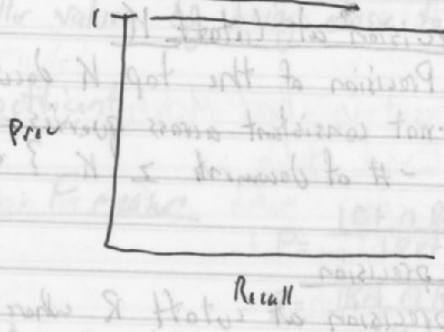


List

Rank	Relevance	Prec	Rec
1	R	(1)	1/5
2	N	1/2	1/5
3	R	(2/3)	2/5
4	N	2/4	2/5
5	N	2/5	2/5
6	R	(3/6)	3/5
7	N	3/7	3/5
8	R	(4/8)	4/5
9	N	4/9	4/5
10	R	(5/10)	5/5



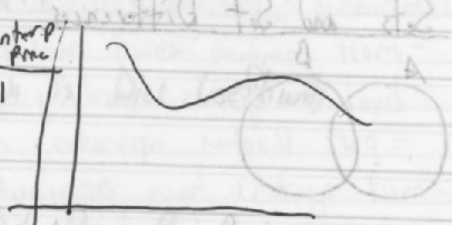
Perfect Retrieval



Area under P-R curve is aprox. Avg Precision

List

Rank	Relevance	Prec	Recall	Interp Prec
1	R	1	1/5	1
2	N	1/2	1/5	
3	N	1/3	1/5	
4	R	2/4	2/5	2/3
5	R	3/5	3/5	2/3
6	R	4/6	4/5	2/3
7	N	4/7	4/5	
8	R	5/8	5/5	1/2
9	N	5/9	5/5	
10	N	5/10	5/5	



want to plot trend (Interpolated Recall-Precision)
- for a given recall, what is the highest precision anywhere beyond that
- Interpolated Precision always drops

Precision at Cutoff K

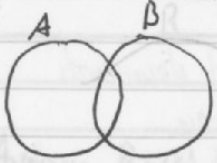
Precision of the top K documents
- not consistent across queries
  - # of documents ≥ K ? < K

R-precision

- precision at cutoff R where R is # of documents for query
- value of 1 if and only if perfect retrieval

5/5/2005

Sets and Set Differences



Jaquard Coeff: $\frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$

Dice Coeff: $\frac{2|A \cap B|}{|A| + |B|}$

Given prec P
rec R

proposition: (arithmetic) average

$M = \frac{P+R}{2}$

Incorrect: b/c for $M = 0.5$

① $P = 0.5$ and $R = 0.5$ doing something interesting
② $P \to 0$ + $R \to 1$ } trivial performance
③ $P \to 1$ + $R \to 0$ }

Geometric Mean: $G = \sqrt{P \cdot R}$ $\quad H \leq G \leq M$

Harmonic Mean: $H = \frac{1}{\frac{1}{2}\left(\frac{1}{P} + \frac{1}{R}\right)} = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2PR}{P+R}$

Harmonic Mean penalizes smaller values by being closer to smaller value
-also similar to Dice Coefficient

Harmonic Mean of P & R: F measure

$$F = \frac{2PR}{P+R}$$

$$= \frac{2 \cdot \frac{|Ret \cap Rel|}{|Ret|} \cdot \frac{|Ret \cap Rel|}{|Rel|}}{\frac{|Ret \cap Rel|}{|Ret|} + \frac{|Ret \cap Rel|}{|Rel|}}$$

$$= \frac{2|Ret \cap Rel|}{|Rel| + |Ret|}$$ which is Dice Coefficient

$$P = \frac{|Ret \cap Rel|}{|Ret|}$$

$$R = \frac{|Ret \cap Rel|}{|Rel|}$$

can also parameterize to weight P or R

$$\frac{1}{\alpha\left(\frac{1}{P}\right) + (1-\alpha)\left(\frac{1}{R}\right)}$$