Evaluation: Precision & Recall

COMP3009J: Information Retrieval

Dr. David Lillis (david.lillis@ucd.ie)

UCD School of Computer Science
Beijing Dublin International College

Precision / Recall

- **Precision** and **Recall** are the two most basic (and most widely used) evaluation metrics in IR, upon which many others are based.
- **Precision** is the fraction of the set of retrieved documents (*Ret*) that are relevant (i.e. that are also in *Rel*)
  - $Precision = \frac{|RelRet|}{|Ret|}$
- **Recall** is the fraction of the relevant documents (*Rel*) that have been retrieved (i.e. they are also in *Ret*).
  - $Recall = \frac{|RelRet|}{|Rel|}$

Example

- Looking at the example, we can see:
  - The number of relevant documents for the query is $|Rel| = 10$
  - The number of retrieved documents for the query is $|Ret| = 15$
  - The number of relevant documents in the retrieved set is $|RelRet| = 5$

- $Precision = \frac{|RelRet|}{|Ret|} = \frac{5}{15} = 0.33$
- $Recall = \frac{|RelRet|}{|Rel|} = \frac{5}{10} = 0.5$

1.	d_{123}	6.	d_9	11.	d_{38}
2.	d_{84}	7.	d_{511}	12.	d_{48}
3.	d_{56}	8.	d_{129}	13.	d_{250}
4.	d_6	9.	d_{187}	14.	d_{113}
5.	d_8	10.	d_{25}	15.	d_3

Rel = {d_3, d_5, d_9, d_{25}, d_{39}, d_{44}, d_{56}, d_{71}, d_{89}, d_{123}}

Precision vs. Recall

- The two metrics of precision and recall are often inversely related: as one increases the other decreases.
- Precision will be high whenever a system is good at avoiding non-relevant documents.
  - A system can achieve very high precision by retrieving very few documents.
- Recall will be high whenever a system finds many relevant documents.
  - A system can achieve 100% recall simply by retrieving all the documents in the collection.

Precision vs. Recall

- Which is most important?
- That is task-dependent!
- Example: Users searching the web want **high precision** (i.e. they want the returned documents to be relevant).
- Because of the size of the web, there are very many documents that will help satisfy the information need, so the user does not need to examine all of them (i.e. recall does not need to be high).
- Instead, the user wishes to avoid wasting time looking at non-relevant documents.

Precision vs. Recall

- On the other hand, a patent lawyer researching a patent must ensure that they get all of the relevant documents and hence they want **high recall**.
- If any relevant documents are missed, this may have serious consequences, so it is essential that all relevant documents are returned.
- They will tolerate lower precision to facilitate this (i.e. they are more likely to be willing to read through some non-relevant documents).
- Ideally every IR system would have both recall and precision of 100% (the answer set is equal to the relevant set).

Precision vs. Recall

- For older, smaller document collections, the calculation of precision and recall was easy.
- Nowadays, because document collections are so huge, it is very difficult to identify all the relevant documents for every query.
- This makes it often impossible to calculate recall accurately.
- Thus, precision tends to be preferred, in addition to other metrics based on precision.

Problems with Precision

- Precision is a **set-based**, **unranked** retrieval metric (as is recall)
- It is a single-value metric based on the entire list of results that was returned by a search engine.
- It does not assess the way the results are **ranked**.
- Most IR systems return ranked lists, not sets of documents, so users are most likely to look at the top of the list first.

Problems with Precision

Rank	Engine A	Engine B
1	N	R
2	N	R
3	N	N
4	R	N
5	R	N

- N = Non-relevant
- R = Relevant
- Precision = 40% for both A and B
- **BUT**, clearly Engine B is better, because its top-ranked documents are relevant, as a user would expect.