Information Retrieval
Evaluation of IR Systems

Bruno Martins
Departamento de Engenharia Informática
Instituto Superior Técnico

1st Semester
2012/2013
1. Evaluation and Relevance
2. Precision vs. Recall
3. Other Measures
4. Reference Collections
IR System Evaluation

- Why evaluate?
  - Measure the benefit of using an IR system
  - Measure how well an IR system fulfills its goal
  - Compare IR systems

- What to evaluate?
  - Collection coverage
  - Processing time
  - Output presentation
  - User effort
  - Recall and Precision
Recall and Precision

Measure the ability of a system to return relevant documents.

Relevance

- Subjective notion
  - Usually evaluated by a set of experts
Outline

1. Evaluation and Relevance
2. Precision vs. Recall
3. Other Measures
4. Reference Collections
Measuring Precision and Recall

Definition

Let $A$ be the set of documents retrieved for query $Q$. Let $R$ be the set of documents that are relevant to query $Q$.

**Precision** is the proportion of retrieved documents that are relevant, i.e.:

$$Pr = \frac{|R \cap A|}{|A|}$$

**Recall** is the proportion of relevant documents retrieved, i.e.:

$$Re = \frac{|R \cap A|}{|R|}$$
Retrieved documents are ordered ⇒ we are interested in measuring how precision changes as recall increases

Example

Let \( A = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}\} \) be an ordered set of retrieved documents, for a query \( Q \). Let \( R = \{d_2, d_5, d_8, d_{15}\} \) be the set of relevant documents for query \( Q \).

<table>
<thead>
<tr>
<th>Re</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>0.50</td>
<td>0.40</td>
</tr>
<tr>
<td>0.75</td>
<td>0.38</td>
</tr>
</tbody>
</table>
Interpolated Precision-Recall

- Precision is usually measured at 10 standard recall points: 0%, 10%, 20%, ..., 90%, 100%
- Precision at $r\%$ recall is defined as

$$P(r) = \max_{i \geq r} P(i)$$

- Precision is zero after no more relevant documents are found
Example

Let $A = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}\}$ be an ordered set of retrieved documents, for a query $Q$. Let $R = \{d_2, d_5, d_8, d_{15}\}$ be the set of relevant documents for query $Q$.

<table>
<thead>
<tr>
<th>$Re$</th>
<th>$Pr$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>0.10</td>
<td>0.50</td>
</tr>
<tr>
<td>0.20</td>
<td>0.50</td>
</tr>
<tr>
<td>0.30</td>
<td>0.40</td>
</tr>
<tr>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>0.50</td>
<td>0.40</td>
</tr>
<tr>
<td>0.60</td>
<td>0.38</td>
</tr>
<tr>
<td>0.70</td>
<td>0.38</td>
</tr>
<tr>
<td>0.80</td>
<td>0.00</td>
</tr>
<tr>
<td>0.90</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Interpolated Precision-Recall (cont.)

Example

Precision vs. Recall

- non-pruned
- lbpm
- carmel

Reference Collections
Outline

1. Evaluation and Relevance
2. Precision vs. Recall
3. Other Measures
4. Reference Collections
Other Performance Measures

- \textbf{P@N} - Precision at the $N$-th retrieved documents. Most commonly used $P@5$, $P@10$ and $P@20$

- \textbf{R-precision} - Precision at the $R$-th document, where $R$ is the number of relevant documents

- \textbf{F-measure} - Harmonic mean of precision and recall:

\[
F_1 = \frac{2 \times Re \times Pr}{Re + Pr}
\]
Other Performance Measures (cont.)

- **AP** - Average of the values for the precision at each recall point
- **MAP** - Mean average precision

\[
MAP = \frac{\sum_{q=1}^{N} AP_q}{N}
\]
Other Performance Measures (cont.)

- **DCG** - Discounted cumulative gain

\[
DCG_L = R_1 + \sum_{i=1}^{L} \frac{R_i}{\log_2(i)}
\]

where \( R_i = 1 \) if document at rank \( i \) is relevant and \( R_i = 0 \) otherwise.

- **nDCG** - Normalized discounted cumulative gain

\[
nDCG_L = \frac{DCG_L}{\text{Ideal DCG}_L}
\]
Other Performance Measures (cont.)

- **MRR** - Mean reciprocal rank

\[
MRR = \frac{1}{N} \sum_{q=1}^{N} \frac{1}{\text{rank}_q}
\]

where \(\text{rank}_q\) is the rank of the first relevant document.
Outline

1. Evaluation and Relevance
2. Precision vs. Recall
3. Other Measures
4. Reference Collections
Reference Collections

TREC Various collections of documents (news, articles, etc.)
TREC-WEB Pages extracted from the web
CACM articles from Communications of the ACM
ISI Information science papers
CFC Cystic Fibrosis Collection

- Standards for research in IR
- Provide sets queries + evaluated documents
Questions?