Discussion Session
Week 9

INF 141: Information Retrieval
Winter 2009

Yasser Ganjisaffar
yganjisa@ics.uci.edu
Outline

• Assignment 06 Questions
• Web Search Evaluation
Failures!

• Task attempt_200903011033_0154_m_000185_0 failed to report status for 600 seconds. Killing!

• org.apache.hadoop.util.DiskChecker$DiskErrorException: Could not find any valid local directory for taskTracker/jobcache/job_200903011033_0154/jars

• java.lang.NoClassDefFoundError: edu/uci/ics/crawler4j/crawler/HTMLParser

• KILLED
How to Evaluate Search Results?
Expert Labeling of Search Results

- Highly Relevant
- Relevant
- Non-relevant
Ideal Ranking of Results
How to Compare Current Ranking with Ideal Ranking?
Cumulative Gain (CG)

<table>
<thead>
<tr>
<th>Type</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Relevant</td>
<td>2</td>
</tr>
<tr>
<td>Relevant</td>
<td>1</td>
</tr>
<tr>
<td>Non-relevant</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
CG_{10} = 4 \times 2 + 3 \times 1 + 3 \times 0 = 11
\]

\[
CG_5 = 2 \times 2 + 1 \times 1 + 1 \times 0 = 5
\]

\[
CG_2 = 1 \times 2 + 1 \times 1 = 3
\]
Cumulative Gain

\[ CG_{10} = 4 \times 2 + 3 \times 1 + 3 \times 0 = 11 \]
Discounted Cumulative Gain

- Assumptions:
  - Highly relevant documents are more useful when appearing earlier in a search engine result list (have higher ranks).
  - Highly relevant documents are more useful than marginally relevant documents, which are in turn more useful than irrelevant documents.

- Measures the *gain* of a document based on its position in the result list.
Discounted Cumulative Gain

$$DCG_p = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2 i}$$

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{\log_2 (1 + i)}$$
DCG Calculations

\[
\frac{2^r(p) - 1}{\log(1 + p)}
\]

1/\log(2) = 1
3/\log(3) = 1.9
3/\log(4) = 1.5
0
1/\log(6) = 0.4
3/\log(7) = 1.1
0
3/\log(9) = 0.9
1/\log(10) = 0.3
0

DCG_{10} = 7.1
# Ideal DCG

![Diagram showing Ideal DCG](image)

<table>
<thead>
<tr>
<th>Position</th>
<th>Relevance ($r(p)$)</th>
<th>DCG Calculation</th>
<th>$2^{r(p)} - 1$</th>
<th>IDCG$_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$r(p)$</td>
<td>$3 / \log(2)$</td>
<td>$3$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$r(p)$</td>
<td>$3 / \log(3)$</td>
<td>$1.9$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$r(p)$</td>
<td>$3 / \log(4)$</td>
<td>$1.5$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$r(p)$</td>
<td>$3 / \log(5)$</td>
<td>$1.3$</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$r(p)$</td>
<td>$1 / \log(6)$</td>
<td>$0.4$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>$r(p)$</td>
<td>$1 / \log(7)$</td>
<td>$0.36$</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>$r(p)$</td>
<td>$1 / \log(8)$</td>
<td>$0.33$</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>$r(p)$</td>
<td>$0$</td>
<td>$0$</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>$r(p)$</td>
<td>$0$</td>
<td>$0$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$r(p)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$8.79$</td>
</tr>
</tbody>
</table>

The IDCG$_{10}$ is calculated using the formula:

$$\frac{2^{r(p)} - 1}{\log(1 + p)}$$

For example:

- For position 1, $2^{r(1)} - 1 = 3$ and $\log(1 + 1) = \log(2) = 1$, so $\frac{3}{1} = 3$.
Normalized DCG

Normalized DCG

\[ \text{nDCG}_p = \frac{DCG_p}{IDCG_p} \]

- \[ 1/\log(2) = 1 \]
- \[ 3/\log(3) = 1.9 \]
- \[ 3/\log(4) = 1.5 \]
- \[ 0 \]
- \[ 1/\log(6) = 0.4 \]
- \[ 3/\log(7) = 1.1 \]
- \[ 0 \]
- \[ 3/\log(9) = 0.9 \]
- \[ 1/\log(10) = 0.3 \]
- \[ 0 \]
Drawback of DCG?

• Labeling results is expensive.
• No ideal ordering of results when only partial relevance feedback (labels) is available.
Click-through Data: Implicit Feedback

1  ... ... ... ... ... ... ... ...
2  ... ... ... ... ... ... ... ...
3  ... ... ... ... ... ... ... ...
4  ... ... ... ... ... ... ... ...
5  ... ... ... ... ... ... ... ...
6  ... ... ... ... ... ... ... ...
7  ... ... ... ... ... ... ... ...
8  ... ... ... ... ... ... ... ...
9  ... ... ... ... ... ... ... ...
10 ... ... ... ... ... ... ... ...

Assuming that user has checked results from top to bottom:

2 is more relevant than 1.
5 is more relevant than 1, 3, 4
7 is more relevant than 1, 3, 4, 6

(2,1)  (5,1)  (5,3)  (5,4)
(7,1)  (7,3)  (7,4)  (7,6)
Learning to Rank

- An ideal search engine should rank “2” higher than “1”.
- We can use this training data to learn how to rank search results.
Learning to Rank

<table>
<thead>
<tr>
<th>TF-IDF1</th>
<th>TF-IDF2</th>
<th>PageRank1</th>
<th>PageRank2</th>
<th>Age1</th>
<th>Age2</th>
<th>Title Score1</th>
<th>Title Score2</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2,1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5,1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Google uses more than 200 features