Implicit User Feedback

Hongning Wang
CS@UVa
Explicit relevance feedback

Query → Retrieval Engine → Document collection → Feedback → Updated query

Results:

\[ \text{d}_1 \ 3.5 \]
\[ \text{d}_2 \ 2.4 \]
\[ ... \]
\[ \text{d}_k \ 0.5 \]
\[ ... \]

User judgment

Judgments:

\[ \text{d}_1 + \]
\[ \text{d}_2 - \]
\[ \text{d}_3 + \]
\[ ... \]
\[ \text{d}_k - \]
\[ ... \]
Relevance feedback in real systems

• Google used to provide such functions

- Relevant
- Nonrelevant

Personalization - Wikipedia, the free encyclopedia

Personalization involves using technology to accommodate the differences between individuals. Once confined mainly to the Web, it is increasingly becoming a ... en.wikipedia.org/wiki/Personalized - 42k - Cached - Similar pages - 🌐

Personalized Gifts from Personalization Mall

It shows you went out of your way to find the perfect gift and to personalize it to make it theirs alone! At PersonalizationMall.com, we design most of our ... www.personalizationmall.com/Default.aspx?did=111028 - 47k - Cached - Similar pages - 🌐

What is personalization? - a definition from Whatis.com

Mar 6, 2007 ... On a Web site, personalization is the process of tailoring pages to individual users' characteristics or preferences. searchcrm.techtarget.com/sDefinition/0,,sid11_gci532341,00.html - 72k - Cached - Similar pages - 🌐

— Vulnerable to spammers
How about using clicks

- Clicked document as relevant, non-clicked as non-relevant
  - Cheap, largely available
Is click reliable?

• Why do we click on the returned document?
  – Title/snippet looks attractive
    • We haven’t read the full text content of the document
  – It was ranked higher
    • Belief bias towards ranking
  – We know it is the answer!
Is click reliable?

• Why do not we click on the returned document?
  – Title/snippet has already provided the answer
    • Instant answers, knowledge graph
  – Extra effort of scrolling down the result page
    • The expected loss is larger than skipping the document
  – We did not see it....

*Can we trust click as relevance feedback?*
Accurately Interpreting Clickthrough Data as Implicit Feedback [Joachims SIGIR’05]

• Eye tracking, click and manual relevance judgment to answer
  – Do users scan the results from top to bottom?
  – How many abstracts do they read before clicking?
  – How does their behavior change, if search results are artificially manipulated?
Which links do users view and click?

- Positional bias

  *Fixations: a spatially stable gaze lasting for approximately 200-300 ms, indicating visual attention*

*Figure 1: Percentage of times an abstract was viewed/clicked depending on the rank of the result. First 5 results are visible without scrolling*
Do users scan links from top to bottom?

**Figure 2:** Mean time of arrival (in number of previous fixations) depending on the rank of the result.

View the top two results within the second or third fixation.
Which links do users evaluate before clicking?

- The lower the click in the ranking, the more abstracts are viewed above the click

<table>
<thead>
<tr>
<th>Viewed Rank</th>
<th>Viewed</th>
<th>Clicked Rank</th>
<th>Clicked</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90.6%</td>
<td>76.2%</td>
<td>73.9%</td>
</tr>
<tr>
<td>2</td>
<td>56.8%</td>
<td>90.5%</td>
<td>82.6%</td>
</tr>
<tr>
<td>3</td>
<td>30.2%</td>
<td>47.6%</td>
<td>95.7%</td>
</tr>
<tr>
<td>4</td>
<td>17.3%</td>
<td>19.0%</td>
<td>47.8%</td>
</tr>
<tr>
<td>5</td>
<td>8.6%</td>
<td>14.3%</td>
<td>21.7%</td>
</tr>
<tr>
<td>6</td>
<td>4.3%</td>
<td>4.8%</td>
<td>8.7%</td>
</tr>
</tbody>
</table>

Table 2: Percentage of times the user viewed an abstract at a particular rank before he clicked on a link at a particular rank.
Does relevance influence user decisions?

• Controlled relevance quality
  – Reverse the ranking from search engine

• Users’ reactions
  – Scan significantly more abstracts than before
  – Less likely to click on the first result
  – Average clicked rank position drops from 2.66 to 4.03
  – Average clicks per query drops from 0.8 to 0.64
Are clicks absolute relevance judgments?

• Position bias
   – Focus on position one and two, equally likely to be viewed

<table>
<thead>
<tr>
<th></th>
<th>$l_1^-, l_2^-$</th>
<th>$l_1^+, l_2^-$</th>
<th>$l_1^-, l_2^+$</th>
<th>$l_1^+, l_2^+$</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{rel}(l_1) &gt; \text{rel}(l_2)$</td>
<td>15</td>
<td>19</td>
<td>1</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>$\text{rel}(l_1) &lt; \text{rel}(l_2)$</td>
<td>11</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>$\text{rel}(l_1) = \text{rel}(l_2)$</td>
<td>19</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>total</td>
<td>45</td>
<td>33</td>
<td>4</td>
<td>3</td>
<td>85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$l_1^-, l_2^-$</th>
<th>$l_1^+, l_2^-$</th>
<th>$l_1^-, l_2^+$</th>
<th>$l_1^+, l_2^+$</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{rel}(l_1) &gt; \text{rel}(l_2)$</td>
<td>11</td>
<td>15</td>
<td>1</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>$\text{rel}(l_1) &lt; \text{rel}(l_2)$</td>
<td>17</td>
<td>10</td>
<td>7</td>
<td>2</td>
<td>36</td>
</tr>
<tr>
<td>$\text{rel}(l_1) = \text{rel}(l_2)$</td>
<td>36</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>total</td>
<td>64</td>
<td>36</td>
<td>11</td>
<td>3</td>
<td>114</td>
</tr>
</tbody>
</table>
Are clicks relative relevance judgments?

• Clicks as pairwise preference statements
  – Given a ranked list and user clicks

\[ l_1^{*(2)} \rightarrow l_2 \rightarrow l_3^{*(1)} \rightarrow l_4 \rightarrow l_5^{*(3)} \rightarrow l_6 \rightarrow l_7 \]

• Click > Skip Above
• Last Click > Skip Above
• Click > Earlier Click
• Last Click > Skip Previous
• Click > Skip Next
Clicks as pairwise preference statements

- Accuracy against manual relevance judgment

<table>
<thead>
<tr>
<th>Explicit Feedback</th>
<th>Phase I “normal”</th>
<th>Phase I “normal”</th>
<th>Phase II “swapped”</th>
<th>Phase II “reversed”</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Strategy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inter-Judge Agreement</td>
<td>89.5</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>82.5</td>
</tr>
<tr>
<td>Click &gt; Skip Above</td>
<td>80.8 ± 3.6</td>
<td>88.0 ± 9.5</td>
<td>79.6 ± 8.9</td>
<td>83.0 ± 6.7</td>
<td>83.1 ± 4.4</td>
</tr>
<tr>
<td>Last Click &gt; Skip Above</td>
<td>83.1 ± 3.8</td>
<td>89.7 ± 9.8</td>
<td>77.9 ± 9.9</td>
<td>84.6 ± 6.9</td>
<td>83.8 ± 4.6</td>
</tr>
<tr>
<td>Click &gt; Earlier Click</td>
<td>67.2 ± 12.3</td>
<td>75.0 ± 25.8</td>
<td>36.8 ± 22.9</td>
<td>28.6 ± 27.5</td>
<td>46.9 ± 13.9</td>
</tr>
<tr>
<td>Click &gt; Skip Previous</td>
<td>82.3 ± 7.3</td>
<td>88.9 ± 24.1</td>
<td>80.0 ± 18.0</td>
<td>79.5 ± 15.4</td>
<td>81.6 ± 9.5</td>
</tr>
<tr>
<td>Click &gt; No Click Next</td>
<td>84.1 ± 4.9</td>
<td>75.6 ± 14.5</td>
<td>66.7 ± 13.1</td>
<td>70.0 ± 15.7</td>
<td>70.4 ± 8.0</td>
</tr>
</tbody>
</table>
How accurately do clicks correspond to explicit judgment of a document?

- Accuracy against manual relevance judgment

<table>
<thead>
<tr>
<th>Explicit Feedback Data Strategy</th>
<th>Pages Phase II all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-Judge Agreement</td>
<td>86.4</td>
</tr>
<tr>
<td>Click &gt; Skip Above</td>
<td>78.2 ± 5.6</td>
</tr>
<tr>
<td>Last Click &gt; Skip Above</td>
<td>80.9 ± 5.1</td>
</tr>
<tr>
<td>Click &gt; Earlier Click</td>
<td>64.3 ± 15.4</td>
</tr>
<tr>
<td>Click &gt; Skip Previous</td>
<td>80.7 ± 9.6</td>
</tr>
<tr>
<td>Click &gt; No Click Next</td>
<td>67.4 ± 8.2</td>
</tr>
</tbody>
</table>
What do we get from this user study?

• Clicks are influenced by the relevance of results
  – Biased by the trust over rank positions
• Clicks as relative preference statement is more accurate
  – Several heuristics to generate the preference pairs
How to utilize such preference pairs?

• Pairwise learning to rank algorithms
  – Will be covered later
An eye tracking study of the effect of target rank on web search [Guan CHI’07]

• Break down of users’ click accuracy
  – Navigational search

![Graph showing the effect of target rank on web search accuracy. The graph indicates that users are more likely to click on the first result when it is ranked higher.](image-url)
An eye tracking study of the effect of target rank on web search [Guan CHI’07]

- Break down of users’ click accuracy
  - Informational search

![Diagram showing eye tracking study results]

First result
Users failed to recognize the target because they did not read it!

- Navigational search
Users did not click because they did not read the results!

- Informational search
Predicting clicks: estimating the click-through rate for new ads [Richardson WWW’07]

• To maximize ad revenue
  \[ E_{ad}[Revenue] = \sum_{ad} p(\text{click}|\text{ad}) \cdot CPC(\text{ad}) \]

• Position-bias is also true in online ads
  – Observed low CTR is not just because of ads’ quality, but also their display positions!

Cost per click: basic business model in search engines
Combat position-bias by explicitly modeling it

- Being clicked is related to its quality and position

\[
p(\text{click} | ad, pos) = p(\text{click} | ad, pos, seen)p(\text{seen} | pos) = p(\text{click} | ad, seen)p(\text{seen} | pos)
\]

Calibrated CTR for ads ranking

Discounting factor

\[
- p(\text{click} = 1 | ad, seen = 0) = 0
\]

\[
- p(\text{click} = 1 | ad, seen = 1) = \frac{1}{1 + \exp(-w^T f_{ad})}
\]

Logistic regression by features of the ad
Parameter estimation

• Discounting factor
  – Approximation: positions being clicked must be seen already
    • \( p(\text{seen}|\text{pos}) \propto \#\text{clicks}_\text{at}_\text{pos} \)

• Calibrated CTR
  – Maximum likelihood for \( w \) with historic clicks
    • \( \hat{w} = \arg\max_w \sum_{ad} \log p(\text{click}|\text{ad}, \text{pos}) \)
Calibrated CTR is more accurate for new ads

- Unfortunately, their evaluation criterion is still based biased clicks in testing set
Click models

• Decompose relevance-driven clicks from position-driven clicks
  – Examine: user reads the displayed result
  – Click: user clicks the displayed result
  – Atomic unit: (query, doc)

\[ P(C = 1|d, i) = \sum_{e \in \{0,1\}} P(C = 1|d, i, E = e)P(E = e|d, i) \]

\[ = P(C = 1|d, E = 1)P(E = 1|i) \]
Cascade Model [Craswell et al. WSDM’08]

- Sequential browsing assumption
  - At each position decides whether to move on
    - \( p(C_i = 1) = p(R_i = 1) \prod_{j=1}^{i-1} (1 - p(R_j = 1)) \)
    - Assuming \( R_i = 1 \to C_i = 1 \)
  - Only one click is allowed on each search result page

*Kind of “Click > Skip Above”*?
User Browsing Model [Dupret et al. SIGIR’08]

- Examination depends on distance to the last click

\[ P(c = 1|u, q, r, d) = \alpha u_q \gamma_{rd} \]

- Attractiveness, determined by query and URL
- Examination, determined by position and distance to last click

EM for parameter estimation

From absolute discount to relative discount

Kind of “Click > Skip Next”?
More accurate prediction of clicks

- Perplexity – randomness of prediction
Dynamic Bayesian Model [Chapelle et al. WWW’09]

• A cascade model
  – Relevance quality:

\[
P(E_{i+1} = 1 | E_i = 1, S_i = 0) = \gamma \\
S_i = 1 \Rightarrow E_{i+1} = 0_u
\]
Accuracy in predicting CTR
Revisit User Click Behaviors

Match my query?
Redundant doc?
Shall I move on?

CS@UVa
Content-Aware Click Modeling [Wang et al. WWW’12]

- Encode dependency within user browsing behaviors via descriptive features

Relevance quality of a document: e.g., ranking features

Chance to further examine the result documents: e.g., position, # clicks, distance to last click

Chance to click on an examined and relevant document: e.g., clicked/skipped content similarity

Relevance quality of a document: e.g., ranking features
Quality of relevance modeling

- Estimated relevance for ranking

(a) P@1 ranking performance under different query frequency categories on the random bucket click set

(b) P@1 ranking performance under different query frequency categories on the normal click set
Understanding user behaviors

• Analyzing factors affecting user clicks

<table>
<thead>
<tr>
<th></th>
<th>( f^R )</th>
<th>( w^R )</th>
<th>( f^C )</th>
<th>( w^C )</th>
<th>( f^E )</th>
<th>( w^E )</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>-0.839</td>
<td>0.007</td>
<td>pos</td>
<td>-1.133</td>
<td>0.149</td>
<td>1.807</td>
</tr>
<tr>
<td>authority</td>
<td>0.007</td>
<td>0.098</td>
<td># click</td>
<td>-0.351</td>
<td>0.335</td>
<td>-0.418</td>
</tr>
<tr>
<td>title match</td>
<td>0.098</td>
<td>abs. match</td>
<td>dis. to last click</td>
<td>-0.445</td>
<td>0.415</td>
<td>0.684</td>
</tr>
<tr>
<td>abs. match</td>
<td>0.167</td>
<td>body match</td>
<td>query length</td>
<td>-3.659</td>
<td>3.707</td>
<td>2.947</td>
</tr>
<tr>
<td>body match</td>
<td>0.020</td>
<td>bias</td>
<td>bias</td>
<td>-4.654</td>
<td>4.405</td>
<td>5.325</td>
</tr>
</tbody>
</table>

\( w^C \) indicates the weight of each factor for the C condition, and \( w^E \) for the E condition.

\( f^R \) and \( f^C \) are functions that represent different factors affecting user clicks.

\( w^R \) and \( w^C \) are weights assigned to these factors based on their impact.

\( w^E \) weights are used for evaluating the effectiveness of the factors in the E condition.
What you should know

• Clicks as implicit relevance feedback
• Position bias
• Heuristics for generating pairwise preferences
• Assumptions and modeling approaches for click models