Web Search Engine Evaluation Using Clickthrough Data and a User Model

G. Dupret  B. Piwowarski  V. Murdock

Query Log Analysis: Social and Technological Challenges
Outline

1. Engine Improvement
   - Context
   - State of the Art

2. User Modelling
   - Statistical Model
   - Parameter Estimation
   - Experiments
   - Discussion
Outline

1. Engine Improvement
   - Context
   - State of the Art

2. User Modelling
   - Statistical Model
   - Parameter Estimation
   - Experiments
   - Discussion
Click-Through Data
Applications

- Ranking improvement
  - re-ranking
  - training set for learning
- Engine Evaluation
- Query Recommendation
- Document & Query Clustering
- etc.
Outline

1. Engine Improvement
   - Context
   - State of the Art

2. User Modelling
   - Statistical Model
   - Parameter Estimation
   - Experiments
   - Discussion
Eye tracking experiments:
- Users look at the results sequentially,
- a document selected later is preferred to the documents not selected before,
- a selected document is always preferred to document directly following it,
- etc.

Ad Hoc methods
Outline

1. Engine Improvement
   - Context
   - State of the Art

2. User Modelling
   - Statistical Model
   - Parameter Estimation
   - Experiments
   - Discussion
Model Hypothesis & Definitions

Definitions:

1. **Consideration**: The user saw the document snippet.
2. **Attractivity**: The document snippet is attractive.

Hypothesis:

1. Users browse the result list sequentially,
2. A user selects a document if its
   - considered
   - attractive
Definitions:
1. **Consideration**: The user saw the document snippet.
2. **Attractivity**: The document snippet is attractive.

Hypothesis:
1. Users browse the result list sequentially,
2. A user selects a document if its
   - considered
   - attractive
Model Hypothesis

1. Attractivity depends on document snippet $u$ and query $q$
2. Consideration depends on the distance to the last selection

$$P(s, a, c|u, q, d) = P(s|a, c)P(c|d)P(a|u, q)$$

Consideration and attractivities are Bernoulli experiments:

$$P(a|u, q) = \alpha_{u,q}^a(1 - \alpha_{u,q})^{1-a}$$
$$P(c|d) = \gamma_d^c(1 - \gamma_d)^{1-c}$$
Outline

1. Engine Improvement
   - Context
   - State of the Art

2. User Modelling
   - Statistical Model
   - Parameter Estimation
   - Experiments
   - Discussion
The posterior is the likelihood \( \times \) the priors:

\[
P(D\{|\alpha_u, q\}, \{\gamma\}) = \prod_{d=1}^{D} \text{Be}(\gamma_d|m, n) \prod_{(u,q)} \text{Be}(\alpha_u, q|a, b)
\]

\[
P(D\{|\alpha_u, q\}, \{\gamma\}) = \prod_{d=1}^{D} \left( \prod_{O_n^i \in S_d^o} \gamma_d \alpha_{u_n, q_n} \prod_{O_n^i \in S_d^o} (1 - \gamma_d \alpha_{u_n, q_n}) \right)
\]

\[
\text{Be}(\alpha_u, q|a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \alpha_{u, q}^{a-1} (1 - \alpha_{u, q})^{b-1}
\]

\[
\text{Be}(\gamma_d|m, n) = \frac{\Gamma(m + n)}{\Gamma(m)\Gamma(n)} \gamma_d^{m-1} (1 - \gamma_d)^{n-1}
\]
Posterior Probability of Attractivities & Perseverances

Variational Approximation

\begin{align*}
\alpha_{u,q|D} & \sim \text{Be}(\alpha_{u,q|a + \hat{S}_{u,q}, b + \sum_{d=1}^{D} \hat{S}_{d,u,q}\bar{\gamma}_d}) \\
\gamma_d|D & \sim \text{Be}(\gamma_d|m + \hat{S}_d, n + \sum_{\{u,q\}} \hat{S}_{d,u,q}\bar{\alpha}_{u,q})
\end{align*}
Users look at the results sequentially,
a document selected later is preferred to the documents not selected before,
a selected document is always preferred to document directly following it,
Outline

1. Engine Improvement
   - Context
   - State of the Art

2. User Modelling
   - Statistical Model
   - Parameter Estimation
   - Experiments
   - Discussion
2,952,871 document / query tuples distributed as follows:

<table>
<thead>
<tr>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>320</td>
<td>460</td>
<td>987.6</td>
<td>840</td>
<td>88,130</td>
</tr>
</tbody>
</table>
# "Naive" Model & Comparison

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>likelihood ratio</th>
<th>std. dev.</th>
<th>attractiveness Kendall $\tau$</th>
<th>popularity Kendall $\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>1.88</td>
<td>0.08</td>
<td>0.07</td>
<td>-0.27</td>
</tr>
<tr>
<td>0.1</td>
<td>10</td>
<td>1.75</td>
<td>0.08</td>
<td>-0.03</td>
<td>-0.28</td>
</tr>
<tr>
<td>0.1</td>
<td>1</td>
<td>2.31</td>
<td>0.09</td>
<td>0.03</td>
<td>-0.26</td>
</tr>
<tr>
<td>0.01</td>
<td>1</td>
<td>1.88</td>
<td>0.08</td>
<td>-0.01</td>
<td>-0.27</td>
</tr>
</tbody>
</table>
Results

attractivity

perseverance

position
distance
Score Definition

\[ R = \sum_q P(q) \sum_o P(o|q) \sum_\sigma P(\sigma|o, q) \bar{a}(\sigma, o, q) \]
Engine Scores Degradation

G. Dupret, V. Murdock, B. Piwowarski

Web Search Engine Evaluation
Mean probability of considering a position

\[ DCG_k = rel_1 + \sum_{i=2}^{k} \frac{rel_i}{\log_2 i} \]
Outline

1. Engine Improvement
   - Context
   - State of the Art

2. User Modelling
   - Statistical Model
   - Parameter Estimation
   - Experiments
   - Discussion
Improvements

1. The number of selections should depend on the position,
2. The number of selections should depend on the query,
3. Perseverance should depend on the number of selections the user is willing to do.
4. etc.
Conclusions

Click-through data analysis

1. potentially improve considerably search engines:
   - ranking using a larger training set
   - ranking reflects user needs
   - ranking improves / adapt with time

2. many potential applications:
   - Engine evaluation and comparison
   - Query recommendation, clustering, etc.

3. (Bayesian) Statistical modeling seems to be particularly well suited to this task.
Questions?
Thank you!