

Web Search Engine Evaluation Using Clickthrough Data and a User Model

G. Dupret B. Piwowski V. Murdock

Query Log Analysis: Social and Technological Challenges

Outline

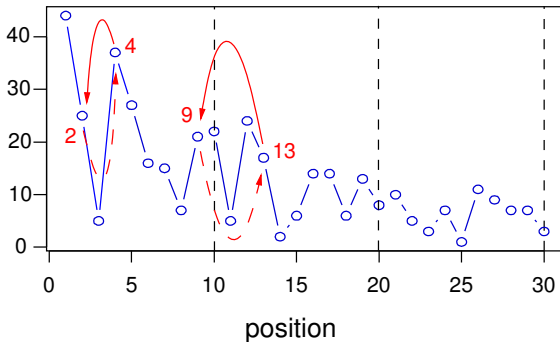
- 1 Engine Improvement
 - Context
 - State of the Art
- 2 User Modelling
 - Statistical Model
 - Parameter Estimation
 - Experiments
 - Discussion

Outline

- 1 Engine Improvement
 - Context
 - State of the Art
- 2 User Modelling
 - Statistical Model
 - Parameter Estimation
 - Experiments
 - Discussion

Click-Through Data

frequency



Applications

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

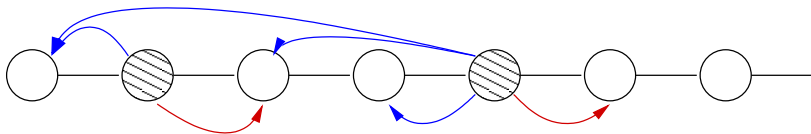
Outline

- 1 Engine Improvement
 - Context
 - State of the Art
- 2 User Modelling
 - Statistical Model
 - Parameter Estimation
 - Experiments
 - Discussion

State of the Art

Relatively new area

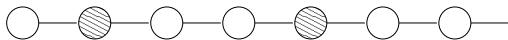
- Eye tracking experiments:
 - Users look at the results sequentially,
 - a document selected later is preferred to the documents not selected before,
 - a selected document is always preferred to document directly following it,
 - etc.
- Ad Hoc methods



Outline

- 1 Engine Improvement
 - Context
 - State of the Art
- 2 **User Modelling**
 - **Statistical Model**
 - Parameter Estimation
 - Experiments
 - Discussion

Model Hypothesis & Definitions



1 Definitions:

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

2 Hypothesis:

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

Model Hypothesis & Definitions



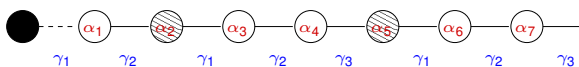
1 Definitions:

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

2 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

Posterior

Summary of the Problem

The posterior is the likelihood \times the priors:

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

$$P(\mathcal{D}|\{\alpha_{u,q}\}, \{\gamma\}) = \prod_{d=1}^D \prod_{\mathcal{O}_n^j \in \dot{\mathcal{S}}_d} \gamma_d \alpha_{u_n, q_n} \prod_{\mathcal{O}_n^j \in \mathring{\mathcal{S}}_d} (1 - \gamma_d \alpha_{u_n, q_n})$$

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

$$\text{Be}(\gamma_d|m, n) = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} \gamma_d^{m-1} (1 - \gamma_d)^{n-1}$$

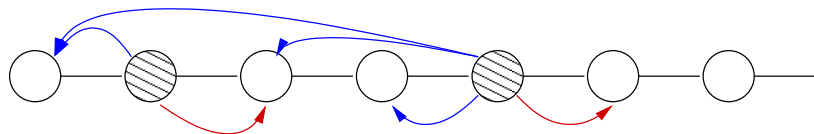
Posterior Probability of Attractivities & Perseverances

Variational Approximation

$$\alpha_{u,q} | \mathcal{D} \sim \text{Be}(\alpha_{u,q} | a + \dot{S}_{u,q}, b + \sum_{d=1}^D \dot{S}_{d,u,q} \bar{\gamma}_d)$$
$$\gamma_d | \mathcal{D} \sim \text{Be}(\gamma_d | m + \dot{S}_d, n + \sum_{\{u,q\}} \dot{S}_{d,u,q} \bar{\alpha}_{u,q})$$

Eye Tracking

- 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

Data

2,952,871 document / query tuples distributed as follows:

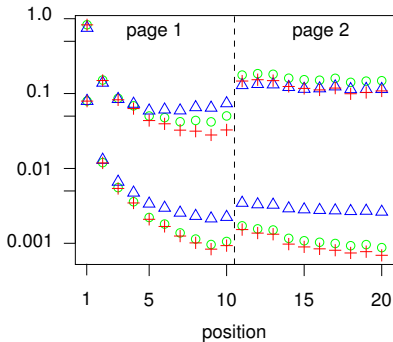
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10	320	460	987.6	840	88,130

"Naive" Model & Comparison

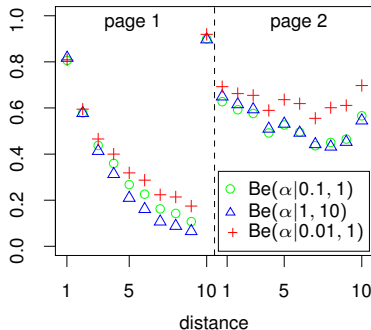
a	b	likelihood ratio	std. dev.	attractivity Kendall $\bar{\tau}$	popularity Kendall $\bar{\tau}$
1	10	1.88	0.08	0.07	-0.27
0.1	10	1.75	0.08	-0.03	-0.28
0.1	1	2.31	0.09	0.03	-0.26
0.01	1	1.88	0.08	-0.01	-0.27

Results

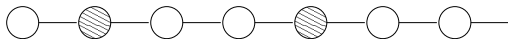
attractivity



perseverance

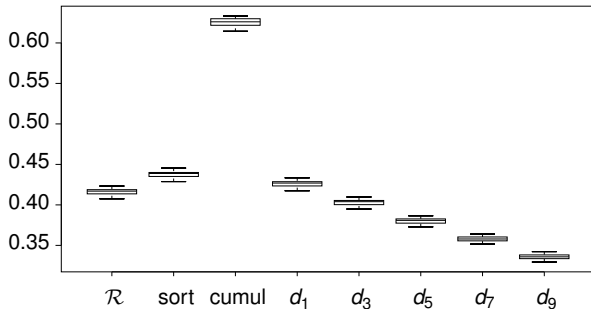


Score Definition



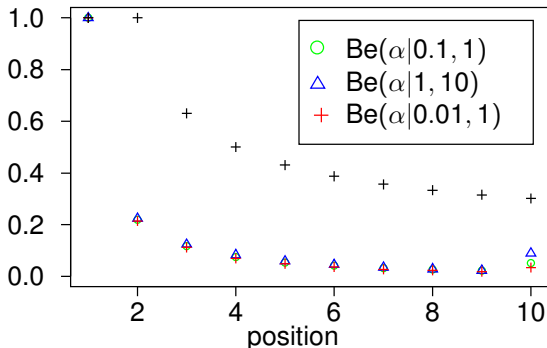
$$\mathcal{R} = \sum_q P(q) \sum_o P(o|q) \sum_\sigma P(\sigma|o, q) \bar{a}(\sigma, o, q)$$

Engine Scores Degradation



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.

End

Questions?

End

Thank you!