Information Retrieval: Personalization

2

Fernando Diaz

Yahoo! Labs

April 25, 2011
Outline

Client-Side Personalization

Server-Side Personalization

Groupization
Client-Side Personalization
Client-Side Personalization

• Advantages

• Disadvantages
Client-Side Personalization

- **Advantages**
  - preserves privacy.
  - richer user signals.

- **Disadvantages**
Client-Side Personalization

- **Advantages**
  - preserves privacy.
  - richer user signals.

- **Disadvantages**
  - no access to data from other users/groups.
  - network latency for interfacing with the index (potentially also subject to query quotas).
  - client space/processing requirements.
Problem Definition

\[ r_t^* = \arg\min_{r \in \mathcal{R}(a_t)} \int_{\mathcal{M}} \ell(a_t, r, \mathbf{m}_t) p(\mathbf{m}_t | U, D, A_t, \mathcal{R}_{t-1}) d\mathbf{m}_t \]

\[ \mathcal{A} = \{\text{query, click, back, next}\} \quad \text{user actions} \]
\[ \mathcal{R}(a) = S_n \quad \text{system rankings} \]
\[ A_t = (a_1, \ldots, a_t) \quad \text{user history} \]
\[ \mathcal{R}_{t-1} = (r_1, \ldots, r_{t-1}) \quad \text{system history} \]
\[ \mathcal{M} \quad \text{all possible user models} \]
Problem Definition

\[ r_t^* = \arg\min_{r \in \mathcal{R}(a_t)} \int_{\mathcal{M}} \ell(a_t, r, m_t) p(m_t | U, D, A_t, \mathcal{R}_{t-1}) dm_t \]

\[ \approx \arg\min_{r \in \mathcal{R}(a_t)} \ell(a_t, r, m_t^*) \]

\[ \mathcal{A} = \{ \text{query, click, back, next} \} \]
\[ \mathcal{R}(a) = S_n \]
\[ \mathcal{A}_t = (a_1, \ldots, a_t) \]
\[ \mathcal{R}_{t-1} = (r_1, \ldots, r_{t-1}) \]
\[ \mathcal{M} \text{ all possible user models} \]
\[ m_t^* = \arg\max_{m_t} p(m_t | U, D, A_t, \mathcal{R}_{t-1}) \]
Representing $m_t$

- Aspects
  - **item identities**: what items has the user indicated are relevant?
    
    $$S \subset \mathcal{D}$$
  
  - **item attributes**: what are the properties of the items the user has indicated are relevant?
    
    $$\mathbf{x} = (x_1, \ldots, x_{|V|})$$
  
  - **demographics**: what are the relevant properties of the user?
    
    not used in this paper

- Composite representation

  $$m_t = (S, \mathbf{x}, A_t, \mathcal{R}_{t-1})$$
Defining $m^*_t$

\[
\begin{align*}
  r^*_t & \approx \argmin_{r \in \mathcal{R}(a_t)} \ell(a_t, r, m^*_t) \\
  & = \argmin_{r \in \mathcal{R}(a_t)} \ell(a_t, r, S, x^*, A_t, \mathcal{R}_{t-1}) \\
  x^* & = \argmax_x p(x|U, D, A_t, \mathcal{R}_{t-1})
\end{align*}
\]
Defining $m_t^*$

$x^* = \alpha q_t + (1 - \alpha) \frac{1}{k} \sum_{i=1}^{t-1} q_i$  
query-based model

$x^* = \alpha q_t + (1 - \alpha) \frac{1}{k} \sum_{s \in S} s$  
snippet-based model
Defining $\ell(a, r, m)$

\[
\ell(a, r, m) = - \sum_{i=1}^{k} p(\text{relevant}|d_i, m) \quad \text{top } k \text{ loss}
\]

\[
\ell(a, r, m) = - \sum_{i=1}^{k} p(\text{view}|d_i)p(\text{relevant}|d_i, m) \quad \text{rank loss}
\]
Constrained Actions

\[ r_t^* \approx \arg\min_r \ell(a_t, r, m) \]
Constrained Actions

\[ r_t^* \approx \arg\min_r \ell(a_t, r, m) = \arg\min_{f(q)} \ell(a_t, f(q), m) \]

\( f(q) \) system ranking for query \( q \)
Constrained Actions

\[ r_t^* \approx \arg\min_r \ell(a_t, r, m) \]
\[ = \arg\min_{f(q)} \ell(a_t, f(q), m) \]
\[ = f(\arg\min_q \ell(a_t, f(q), m)) \]

\[ f(q) \] system ranking for query \( q \)
Implementation

Figure 1: UCAIR architecture

Search Engine (e.g., Google) -> Query Modification

User Modelering

Result Re-Ranking

Result Buffer

Search History Log (e.g., past queries, clicked results)

User

query

clickthrough...

results
Results

<table>
<thead>
<tr>
<th>Ranking Method</th>
<th>prec@5</th>
<th>prec@10</th>
<th>prec@20</th>
<th>prec@30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>0.538</td>
<td>0.472</td>
<td>0.377</td>
<td>0.308</td>
</tr>
<tr>
<td>UCAIR</td>
<td>0.581</td>
<td>0.556</td>
<td>0.453</td>
<td>0.375</td>
</tr>
<tr>
<td>Improvement</td>
<td>8.0%</td>
<td>17.8%</td>
<td>20.2%</td>
<td>21.8%</td>
</tr>
</tbody>
</table>
Results

Figure 5: Precision at top 20 documents of UCAIR and Google

Figure 6: Precision at top 20 result of UCAIR and Google
Server-Side Personalization
Server-Side Personalization

- Client-side personalization is inherently decentralized.
  - Evaluation requires a controlled experiment similar to a user study.
  - Testing new systems/algorithms difficult.
- Server-side personalization centralizes data collection, allowing in situ evaluation.
Server-Side Evaluation

Average Rank

\[ \text{average rank}_{q,u} = \frac{1}{\sum_i \delta_{q,u}(i)} \sum_i \delta_{q,u}(i) \times i \]

\[ \delta_{q,u}(i) \text{ document at rank } i \text{ was clicked for query } q \text{ by user } u \]
Average Rank

\[ \text{average rank}_{q,u} = \frac{1}{\sum_i \delta_{q,u}(i)} \sum_i \delta_{q,u}(i) \times i \]

\( \delta_{q,u}(i) \) document at rank \( i \) was clicked for query \( q \) by user \( u \)

What’s the problem with this metric?
Server-Side Evaluation

Rank Scoring

\[ R_{q,u} = \sum_{i=1}^{k} \frac{\delta_{q,u}(j)}{2(j-1)/(\alpha-1)} \]

What metric is this a version of?
Server-Side Evaluation

Rank Scoring

\[ R_{q,u} = \sum_{i=1}^{k} \frac{\delta_{q,u}(j)}{2^{(j-1)/(\alpha-1)}} \]

\[ \text{gain} = \delta_{q,u}(j) \]

\[ \text{discount} = 2^{(j-1)/(1-\alpha)} \]

What metric is this a version of?
Server-Side Evaluation

Rank Scoring

\[(\log_2(r + 1))^{-1}\]

\[2^{(r-1)/(1-\alpha)}, \alpha = 5\]
Server-Side Personalization

P-Click

\[ f_{q,u}^p(p) = \frac{\delta_{q,u}(p)}{\beta + \sum_p \delta_{q,u}(p)} \]
Server-Side Personalization

P-Click

\[ f_{q,u}^p(p) = \frac{\delta_{q,u}(p)}{\beta + \sum_p \delta_{q,u}(p)} \]

- position bias
- no generalization to new pages
Server-Side Personalization

L-Profile

\[ f_{q,u}^L(p) = \frac{\sum_c c(u) \times c(p)}{\sqrt{\sum_c c(u)^2} \times \sqrt{\sum_c c(p)^2}} \]

c(u) category distribution for user u

c(p) category distribution for page p
Server-Side Personalization

L-Profile

c(u) = \sum_p P(p|u)w(p)c(p)

P(p|u) = \frac{\sum_q \delta_{q,u}(p)}{\sum_{q,p} \delta_{q,u}(p)}

w(p) = \log \frac{|\mathcal{U}|}{|\mathcal{U}(p)|}

What effect does \( w \) have?
Server-Side Evaluation

L-Profile

$\log \frac{|U|}{|U(p)|}$, $|U| = 100$
Server-Side Personalization

short-term models

• S-Profile
  • Oftentimes, user information needs are short-term.
  • We can also define a profile based on a recent subset of pages and categories.

• LS-Profile

\[ f_{q,u}^{LS}(p) = \lambda f_{q,u}^{L}(p) + (1 - \lambda) f_{q,u}^{S}(p) \]
Server-Side Personalization

G-Click

\[ f_{q,u}^G(p) = \frac{\sum_{v \in U} \langle u, v \rangle \delta_{q,v}(p)}{\sum_{v \in U, p'} \langle u, v \rangle \delta_{q,v}(p')} \]

\[ \langle u, v \rangle = \frac{\sum_c c(u) \times c(v)}{\sqrt{\sum_c c(u)^2} \times \sqrt{\sum_c c(v)^2}} \]
Server-Side Personalization

Experiments

• try to improve baseline ranking algorithm
• parameters set empirically (on test set?)
  • may reflect upper bound on performance.
  • be careful about generalizing these results.
• Baseline is original web ranking (WEB)
  • what other naïve baseline is missing?
Server-Side Personalization

Results

<table>
<thead>
<tr>
<th>method</th>
<th>all</th>
<th>not-optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R.S.</td>
<td>A.R.</td>
</tr>
<tr>
<td>WEB</td>
<td>69.4669</td>
<td>3.9240</td>
</tr>
<tr>
<td>P-Click</td>
<td>70.4350</td>
<td>3.7338</td>
</tr>
<tr>
<td>L-Profile</td>
<td>66.7378</td>
<td>4.5466</td>
</tr>
<tr>
<td>S-Profile</td>
<td>66.7822</td>
<td>4.4244</td>
</tr>
<tr>
<td>LS-Profile</td>
<td>68.5958</td>
<td>4.1322</td>
</tr>
<tr>
<td>G-Click</td>
<td>70.4168</td>
<td>3.7361</td>
</tr>
</tbody>
</table>

- *-Click > WEB
- P-Click ≈ G-Click
- *-Profile < WEB
Server-Side Personalization

Results

![Graphs showing distribution of rank scoring increments over WEB method]

Figure 5: Distributions of rank scoring increment over WEB method.

Figure 6: Search accuracy improvements over WEB method on the same rank scoring increment range is plotted in the y-axis with log scale.

Table 3(a) shows the personalization performance on the different search frequencies in the Figure (b), smaller average means better performance.

Table 3(b) gives the results on the different search frequencies in Figure (b), smaller average means better performance.

In Subsection 5.4, we find about 46% test queries are repeated non-optimal queries repeated by either the same user or different users and 33% queries have greater search activities in training days do not become more unstable when the user has more and more search activities in training days.

This is because there is more noise in queries and further more analysis on user's real information need and select only the most related results of rank scorings on queries given by users with different search frequencies in Figure 7. We find:

1. Method P-Click when using a user's long-term interests gets a better performance.
2. Method L-Profile when using a user's long-term interests performs similar averages for users with different search frequencies.
3. Method G-Click is very close to each other.

The high repetition ratio in real-world makes these repeated non-optimal queries repeated by either the same user or different users.

These results tell us that we should record user clicks. The high repetition ratio in real-world makes these repeated non-optimal queries repeated by either the same user or different users.

This is because we can catch users' long-term interests more accurately when their search histories are long enough. At the same time, the personalization based on his/her past clicks performs well. These results tell us that we should record user clicks.

These results tell us that we should record user clicks.

Methods P-Click and G-Click are based on historical forms well. These results tell us that we should record user clicks.

From personalized search? Do profile-based personalized search methods they proposed is achieved when appropriate search histories to build up user profiles.

Tan et al. [31] find that the best performance of profile-based personalized search methods they proposed is achieved when appropriate search histories to build up user profiles.

 distintive ways for users to review their search histories, and the personalization based on his/her past clicks performs well. These results tell us that we should record user clicks.

Forms well. These results tell us that we should record user clicks.
Do users who frequently use search engines benefit more from personalized search? Just like those provided by some current search engines, convenient ways for users to review their search histories, search if no privacy problems exist. We also should provide query and click histories and use them to improve future forms well. These results tell us that we should record user and the personalization based on his/her past clicks per-same user. This means that when a user re-submits a query, personalization methods $P$-Click and $G$-Click have significant the queries repeated by the same user. We find the per-user or di-repeated non-optimal queries repeated by either the same user or di-clicks. The high repetition ratio in real-world makes these click-based personalization strategies work well. $P$-Click and $G$-Click are based on historical based on a user's previous queries and clicks. In this pa-

### 6.3 Performance on Repeated queries

*Figure 6: Search accuracy improvements over WEB method on the same rank scoring increment range is plotted in y-axis with log scale.*

*Figure 5: Distributions of rank scoring increment over WEB method.*

In Subsection 5.4, we find about 46% test queries are re-

---

### Table 3(a) shows the personalization performance on the different users and Table 3(b) gives the results on

<table>
<thead>
<tr>
<th>Number of test queries</th>
<th>$P$-Click</th>
<th>WEB</th>
<th>$L$-Profile</th>
<th>$P$-Click - WEB</th>
<th>$L$-Profile - WEB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

---

### 33% have also observed that re-finding behavior is common, especially when the number of queries grows. To answer these questions, we plot the improve-

*Figure 7: Repeated query accuracy for different profile-based methods.*
In this paper, we try to investigate whether personalized search can improve the search performance. We use the historical query logs of MSN to evaluate five personalized search strategies. The experimental results show that the profile-based personalized search strategies perform badly. Unlike the click entropy, the personalization strategies perform less optimally, which is due to the rough implementation of our strategies. The personalization strategies are straightforward and do not need more reliable. The appropriate combination of them can be effective under different query situations. We find all proposed methods have significant improvement. In other words, both long-term and short-term context can gain better performance than solely using either of them. In other words, both long-term and short-term search contexts are very important to personalize search results. The combination of the two type of search context can make the prediction of real user information more valid and secure.

Table 3: Performance on repeated queries. In Table (a), Y means that the query is repeated by either the same user or different query frequencies.

<table>
<thead>
<tr>
<th>Method</th>
<th>WEB</th>
<th>P-Click</th>
<th>G-Click</th>
<th>L-Profile</th>
<th>LS-Profile</th>
<th>S-Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>WEB</td>
<td>0%</td>
<td>-5%</td>
<td>-10%</td>
<td>0%</td>
<td>5%</td>
<td>-5%</td>
</tr>
<tr>
<td>P-Click</td>
<td>-5%</td>
<td>0%</td>
<td>-5%</td>
<td>-5%</td>
<td>0%</td>
<td>-5%</td>
</tr>
<tr>
<td>G-Click</td>
<td>-10%</td>
<td>-5%</td>
<td>0%</td>
<td>-5%</td>
<td>-5%</td>
<td>0%</td>
</tr>
<tr>
<td>L-Profile</td>
<td>0%</td>
<td>-5%</td>
<td>-5%</td>
<td>-10%</td>
<td>-10%</td>
<td>-5%</td>
</tr>
<tr>
<td>LS-Profile</td>
<td>5%</td>
<td>-5%</td>
<td>-10%</td>
<td>5%</td>
<td>-10%</td>
<td>5%</td>
</tr>
<tr>
<td>S-Profile</td>
<td>-5%</td>
<td>0%</td>
<td>-5%</td>
<td>-5%</td>
<td>0%</td>
<td>-5%</td>
</tr>
</tbody>
</table>

Figure 7: Rank scoring increments over WEB for different query frequencies.
Groupization
Types of Groups

- Task
- Age
- Job team
- Gender
- Job role
- Location
- Interest group
- Relevance judgments
- Query selection
- Desktop content
- Task-based
- Trait-based
- Group Longevity

Group Identification

Explicit

Implicit
Groupization

- Advantages
  - Can preserve (group) privacy.
  - Potentially richer user signals.

- Disadvantages
  - No access to data from external groups.
  - Client space/processing requirements.
Groupization

- **Advantages**
  - can preserve (group) privacy.
  - potentially richer user signals.

- **Disadvantages**
Groupization

- Advantages
  - can preserve (group) privacy.
  - potentially richer user signals.

- Disadvantages
  - no access to data from external groups.
  - client space/processing requirements.
Experiments

1. Use group-level information instead of strictly personal information for personalization.
2. Groups derived from social and work interests.
3. Users sampled from a work-related pool.
how much of this is based on the sampled users?
Summary

1. Many (related) algorithms for personalization.
2. Experimental results for personalization—even at a group level—are not conclusive.
3. The one result we can have confidence in: having a lot of data is a prerequisite.
4. Need more published work from large scale production systems working in personalization.