User Modeling & Search Personalization

Jin-young Kim
Agenda

• Part I : Overview & Past
  • [Teevan05]

• Part II : Present & Future
  • [Teevan08] [Teevan09]
Overview
Personalization in Action

- Google ‘Rosie Jones’, before and after log-in
Motivation

• Background
  • Need to satisfy diverse information needs
  • More user data is available

• Benefit
  • More relevant result
  • Less effort on user side
User Modeling for Query Modeling

• User modeling may help
  • When idiosyncrasy prevails
  • e.g. query-term disambiguation / expansion

• User modeling may not help
  • When universal criteria exists
  • e.g. spellchecking / segmentation
Classes of User Model

• Individual vs. Group (granularity)
  • Specificity vs. Data Richness trade-off
• Short-term vs. Long-term (time-frame)
• Behavior-based vs. Profile-based (source)
  • Profile : desktop documents, browsing history
  • Behavior : past queries and clicks
Components of Personalization

1. Pre-retrieval
2. On-retrieval
3. Post-retrieval

User Model

Query

Expanded Query

Universal Ranking

Personalized Ranking

Hybrid Ranking

Final Ranking

User

System
Components of Personalization

- Pre-retrieval
  - Query expansion
- On-retrieval
  - Personalizable query classification
  - Rank-fusion with universal ranking
- Post-retrieval
  - Local re-ranking
Challenges in Personalization

- Sparse data for user modeling
- Groupizing instead of personalizing [Teevan09]
- Not all queries benefit from personalization
- Query classification [Dou07] [Teevan08]
- Efficiency concern
- Client-side implementation
Challenges in Personalization

• Evaluation of Personalization
  • Difficult to use past TREC data
    • Key lies in modeling individual user
  • Difficult to use usual click-log data
    • Requires long-term data with user identification
  • Typically involves user study
    • Collect personal information for user modeling
    • Evaluate with personalized queries and judgments
Related Topics

• Contextual IR
  • Use short-term (session-level) evidence

• Query Classification
  • Classifying personalizable query is important
  • Query intent classification requires user model
Related Topics

• Personal Information Management (PIM)
  • Different aspect of personalization
    • Algorithm vs. Data
  • PIM can help user modeling
  • PIM can benefit from personalization
    • Rich user context
    • Rich training data
Personalizing Search via Automated Analysis of Interests and Activities

- **User Model**
  - Query history & web documents index
  - Desktop index (recent vs. full)

- **Retrieval Model**
  - Relevance feedback with BM25
  - Re-ranking on client side
Personalizing Search via Automated Analysis of Interests and Activities

- **Personal Profile Feedback**

- **Term weight (IDF)**

\[
\begin{align*}
    w_i &= \log \frac{(r_i + 0.5)(N-n_i+0.5)}{(n_i + 0.5)(R-r_i+0.5)} \\
    &= \text{IDF}_{\text{col}} - \text{IDF}_{\text{user}}
\end{align*}
\]
Personalizing Search via Automated Analysis of Interests and Activities

- Experiment
  - Rich user model helps (full index)
  - Title & snippet is enough for corpus/document
  - Personalization is better than ideal RF!
Personalizing Search via Automated Analysis of Interests and Activities

• Conclusion
  • Profile-based personalization helps web search
  • More effective for short & ambiguous queries

• Caveat
  • Small number of queries (130)
A Large-scale Evaluation and Analysis of Personalized Search Strategies

• User Model
  • Based on click history
    • Boost previous clicks of the user\(^P\)
    • Boost previous clicks of the group\(^G\)
  • Based on profile (browsing history)
    • Similarity between page & user vector
    • Short-term\(^S\) / Long-term\(^L\) / Interpolation\(^{LS}\)

• Retrieval Model
  • Re-ranking of top 50 rank list
  • Rank-fusion of web and personalized result
A Large-scale Evaluation and Analysis of Personalized Search Strategies

- **Experiment**
  - Based on 10,000 users with 56,000 queries
  - Click-based methods help, while profile-based methods do not
  - Profile-based methods hurt more queries than it helps

<table>
<thead>
<tr>
<th>method</th>
<th>all</th>
<th></th>
<th>not-optimal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R.S.</td>
<td>A.R.</td>
<td>R.S.</td>
<td>A.R.</td>
</tr>
<tr>
<td>WEB</td>
<td>69.4669</td>
<td>3.9240</td>
<td>47.2623</td>
<td>7.7879</td>
</tr>
<tr>
<td>P-Click</td>
<td><strong>70.4350</strong></td>
<td><strong>3.7338</strong></td>
<td><strong>49.0051</strong></td>
<td><strong>7.3380</strong></td>
</tr>
<tr>
<td>L-Profile</td>
<td>66.7378</td>
<td>4.5466</td>
<td>45.8485</td>
<td>8.3861</td>
</tr>
<tr>
<td>S-Profile</td>
<td>66.7822</td>
<td>4.4244</td>
<td>45.1679</td>
<td>8.3222</td>
</tr>
<tr>
<td>LS-Profile</td>
<td>68.5958</td>
<td>4.1322</td>
<td>46.6518</td>
<td>8.0445</td>
</tr>
<tr>
<td>G-Click</td>
<td>70.4168</td>
<td>3.7361</td>
<td>48.9728</td>
<td>7.3433</td>
</tr>
</tbody>
</table>

(a) P-Click  
(b) L-Profile
A Large-scale Evaluation and Analysis of Personalized Search Strategies

- **Experiment**
  - Personalization helps queries with high click-entropy
  - Profile-based methods benefit from user information (?)
  - Combining short and long term profile improves stability
A Large-scale Evaluation and Analysis of Personalized Search Strategies

- Conclusion
  - Click-entropy indicates personalizability
  - Click-based user model consistently helps

- Caveat
  - Click-based model is hard to generalize
  - Profile-based model is naive
  - No effort to eliminate bias of click log
Follow-up Papers

• [shen05]
  • Query expansion using recent & frequent query-terms
  • Re-ranking based on vector similarity with user profile

• [chirita07]
  • Query expansion using desktop index
  • Adaptive to predicted ambiguity of query
Present
Previous Works Revealed...

- Personalization is beneficial
  - Both user profile and behavior can be used
  - Having a rich user profile is critical
- Knowing when to personalize matters
  - Query ambiguity seems to be the key
To Personalize or Not to Personalize

- Measures of Query Ambiguity
- Potential for personalization
- Inter-rater reliability (Fleiss’ kappa)
- Click entropy
- Predictive Features of Query Ambiguity
- Result entropy

<table>
<thead>
<tr>
<th>Information</th>
<th>Query</th>
<th>History</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query clarity</td>
<td>ODP category entropy</td>
<td>Refomulation probability</td>
<td>Avg/σ click position</td>
</tr>
<tr>
<td># of ODP categories</td>
<td># of distinct ODP categories</td>
<td># of times query issued</td>
<td>Avg/σ seconds to click</td>
</tr>
<tr>
<td># of distinct ODP categories</td>
<td># of URLs matching ODP</td>
<td># of users who issued query</td>
<td>Avg/σ clicks per user</td>
</tr>
<tr>
<td>Portion of results non-html</td>
<td>Portion that are “.com”/”.edu”</td>
<td>Avg/σ # of query suggest.</td>
<td>Click entropy</td>
</tr>
<tr>
<td># of distinct domains</td>
<td># of query suggestions offered</td>
<td>Avg/σ # of ads</td>
<td>Potential for personalization</td>
</tr>
</tbody>
</table>

- Measures of Query Ambiguity
- Predictive Features of Query Ambiguity

- Measures of Query Ambiguity
- Predictive Features of Query Ambiguity
To Personalize or Not to Personalize

• Understanding Ambiguity

• Based on 2.4M query logs and judgments by 128 people

• Explicit & implicit measures correlate

• Low RE queries are hard to personalize

<table>
<thead>
<tr>
<th></th>
<th>Click entropy</th>
<th>Potential at 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Low RE</td>
</tr>
<tr>
<td>Query length (words)</td>
<td>0.20</td>
<td>0.16</td>
</tr>
<tr>
<td>Query length (chars)</td>
<td>-0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>URL fragment</td>
<td>-0.36</td>
<td>-0.23</td>
</tr>
<tr>
<td>Location mentioned</td>
<td>-0.03</td>
<td>-0.04</td>
</tr>
<tr>
<td>Advanced query</td>
<td>-0.01</td>
<td>-0.02</td>
</tr>
<tr>
<td># of query suggestions</td>
<td>0.12</td>
<td>0.15</td>
</tr>
<tr>
<td># of times issued</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td># of distinct users</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Avg. # of results</td>
<td>0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td>% issued during work</td>
<td>-0.10</td>
<td>-0.04</td>
</tr>
<tr>
<td>Query clarity</td>
<td>0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td>Category entropy</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td># of distinct categories</td>
<td>0.01</td>
<td>-0.02</td>
</tr>
<tr>
<td># of URLs in ODP</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>Top level domain entropy</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td># of distinct hosts</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td>Click entropy</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Potential at 10</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>Result entropy (RE)</td>
<td>0.53</td>
<td>-0.04</td>
</tr>
<tr>
<td>Avg. clicks per user</td>
<td>0.73</td>
<td>0.69</td>
</tr>
<tr>
<td>Avg. click position</td>
<td>0.90</td>
<td>0.86</td>
</tr>
<tr>
<td>Avg. seconds to click</td>
<td>0.03</td>
<td>0.05</td>
</tr>
</tbody>
</table>
To Personalize or Not to Personalize

- Predicting Ambiguity
- Result set < Query history < Click history
- More challenging for low RE queries

Table 4. The model accuracy using different features and predicting different targets.

<table>
<thead>
<tr>
<th>Features used</th>
<th>Result entropy</th>
<th>Click entropy Baseline</th>
<th>Accuracy</th>
<th>Potential at 5 Baseline</th>
<th>Accuracy</th>
<th>Potential at 10 Baseline</th>
<th>Accuracy</th>
<th>Potential cluster Baseline</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>All</td>
<td>0.254</td>
<td>0.399</td>
<td>0.256</td>
<td>0.385</td>
<td>0.498</td>
<td>0.498</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>All</td>
<td>0.254</td>
<td>0.399</td>
<td>0.256</td>
<td>0.393</td>
<td>0.498</td>
<td>0.498</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>All</td>
<td>0.254</td>
<td>0.426</td>
<td>0.256</td>
<td>0.389</td>
<td>0.498</td>
<td>0.495</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>All</td>
<td>0.254</td>
<td>0.813</td>
<td>0.256</td>
<td>0.820</td>
<td>0.498</td>
<td>0.611</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Low</td>
<td>0.258</td>
<td>0.360</td>
<td>0.258</td>
<td>0.360</td>
<td>0.342</td>
<td>0.342</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Low</td>
<td>0.258</td>
<td>0.366</td>
<td>0.258</td>
<td>0.357</td>
<td>0.342</td>
<td>0.340</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Low</td>
<td>0.258</td>
<td>0.360</td>
<td>0.258</td>
<td>0.390</td>
<td>0.342</td>
<td>0.341</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Low</td>
<td>0.258</td>
<td>0.788</td>
<td>0.258</td>
<td>0.794</td>
<td>0.342</td>
<td>0.495</td>
</tr>
</tbody>
</table>
To Personalize or Not to Personalize

• Conclusion
  • User interaction history indicates potential benefit for personalization

• Caveat
  • Most effective features require click history
Discovering and Using Groups to Improve Personalized Search

• Motivation
  • Lack of rich user profile is a challenge for personalization
  • User profile can be augmented by grouping

• Problem
  • What is most useful groups for that purpose?
  • How can we groupize web search results?
Discovering and Using Groups to Improve Personalized Search

- Explicit trait-based Groups (long-term)
  - gender, age, job role, ...
- Explicit task-based Groups (short-term)
  - Shared work-related tasks
- Implicit Group
  - Similarity in personal info. and behavior
Discovering and Using Groups to Improve Personalized Search

- Variation Within Groups
  - Query Selection
    - Cohesive within task-based groups
  - User Profile
    - Cohesive within interest-based groups
- Relevance Judgments
  - Only slight agreement in overall (0.08)
  - Task-based groups showed higher agreement (0.16)
Discovering and Using Groups to Improve Personalized Search

- **User Model**
  - Content(profile)-based \cite{Teevan05}
  - Behavior-based \cite{Dou07}

- **Retrieval Model**
  - Balance content & behavior-based algorithms
  - Web ranking was used as prior
Discovering and Using Groups to Improve Personalized Search

- **Groupization Result**
  - All participants (0.63 > 0.60 by personalization)
  - Explicit, trait-based group (0.61 > 0.59)
  - Explicit, task-based group (0.67)
  - Implicit group
Discovering and Using Groups to Improve Personalized Search

• Conclusion
  • Rich user model by grouping helps web search
  • Task-based grouping is particularly helpful
  • Implicit grouping can predict explicit groups

• Caveat
  • Task-based groups are hard to find in general
Future
What have we learned about personalization?

• Many was proposed
  • User Model
  • Retrieval Model
  • Query classification

• None have been evaluated comparatively!
  • Inconsistency among conclusions
Opportunities Ahead

• Synthesis of Methods
  • Profile + behavior-based user model
  • Query expansion + re-ranking
  • Personalizable query classification

• Realistic evaluation
  • Long-term study with many users
  • Qualitative as well as quantitative analysis
Long-term Vision

- Personalization as a Component
  - Combine with localization and groupization
  - Other query processing technique
- Proactive Information Retrieval
  - Some early results
My Work
LiFiDeA Project

- Collect all the personal information
  - Desktop / Web Documents
  - Desktop / Web Activities
- Provide multiple access path
  - Link-based Browsing
  - Full-text Search
Personalization in LiFiDeA

- By Content-type
  - Personalized search of personal information
  - Personalized search of web

- By Evidence
  - Profile-based personalization
  - Context-based personalization