Search Personalization

Jaime Teevan
Microsoft Research

Personalization and Search

- Measuring the value of personalization
  - Do people’s notions of relevance vary?
- Understanding the individual
  - How can we model a person’s interests?
- Calculating personal relevance
  - How can we use the model to measure relevance?
- Other ways to personalize search
  - What other aspects can we personalize?

Personalization and Search

- Measuring the value of personalization
  - An example
    - Lots of relevant results ranked low
    - Best group ranking v. individual ranking
- Understanding the individual
- Calculating personal relevance
- Other ways to personalize search

Relevant Content Ranked Low
Overview

- Measuring the value of personalization
- Understanding the individual
  - Explicit v. implicit
  - Client-side v. server-side
  - Individual v. group
- Calculating personal relevance
- Other ways to personalize search

Learning More Explicitly v. Implicitly

- Explicit
  - User shares more about query intent
  - User shares more about interests
  - Hard to express interests explicitly

Query Words

Washington or Wisconsin?
Undergrad or grad?
Learning More Explicitly v. Implicitly

• Explicit
  – User shares more about query intent
  – User shares more about interests
  – Hard to express interests explicitly

<table>
<thead>
<tr>
<th>Explicit</th>
<th>Implicit</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Intellectual property?</td>
<td>– Hard to express interests explicitly</td>
</tr>
<tr>
<td>– Rock climbing?</td>
<td>– Query context inferred</td>
</tr>
<tr>
<td>– Tobacco and guns</td>
<td>– Profile inferred about the user</td>
</tr>
<tr>
<td>– Arts and Business</td>
<td>– Less accurate, needs lots of data</td>
</tr>
</tbody>
</table>

Profile Information

• Behavior-based
  – Click-through
  – Personal PageRank

• Content-based
  – Categories
  – Term vector

[topic: computers]

Server-side v. Client-side Profile

• Server-side
  – Pros: Access to rich Web/group information
  – Cons: Personal data stored by someone else

• Client-side
  – Pros: Privacy
  – Cons: Need to approximate Web statistics

• Hybrid solutions
  – Server sends necessary Web statistics
  – Client sends some profile information to server

Match Individual to Group

• Can use groups of people to get more data
Match Individual to Group

- Can use groups of people to get more data
- Back off from individual → group → all
- Collaborative filtering

Overview

- Measuring the value of personalization
- Understanding the individual
- Calculating personal relevance
  - Behavior-based example
  - Content-based example
- Other ways to personalize search

Behavior-Based Relevance

- People often want to re-find
- People have trusted sites
- Boost previously viewed URLs or domains

<table>
<thead>
<tr>
<th>Repeat Query</th>
<th>Repeat Click</th>
<th>New Click</th>
</tr>
</thead>
<tbody>
<tr>
<td>33%</td>
<td>28%</td>
<td>4%</td>
</tr>
<tr>
<td>New Query</td>
<td>07%</td>
<td>04%</td>
</tr>
<tr>
<td>20%</td>
<td>61%</td>
<td></td>
</tr>
</tbody>
</table>

Behavior-Based Relevance

- People often want to re-find
- People have trusted sites
- Boost previously viewed URLs or domains

Content-Based Relevance

- Explicit relevance feedback
  - Mark documents relevant
  - Used to re-weight term frequencies

Behavior-Based Relevance

- People often want to re-find
- People have trusted sites
- Boost previously viewed URLs or domains

Behavior-Based Relevance

- People often want to re-find
- People have trusted sites
- Boost previously viewed URLs or domains
Content-Based Relevance

Score = \( \sum tf_i \cdot w_i \)

where:

\[ w_i = \log \left( \frac{N}{n_i} \right) \]

\[ w_i = \log \left( \frac{(r_i+0.5)(N-n_i-R+r_i+0.5)}{(n_i-r_i+0.5)(R-r_i+0.5)} \right) \]

Content-Based Relevance

• Explicit relevance feedback
  – Mark documents relevant
  – Used to re-weight term frequencies

• Lots of information about the user
  – Consider read documents relevant
  – Use to re-weight term frequencies

Personalization Performance

• Personalized search hard to evaluate
• Mostly small improvements despite big gap
• Identify ambiguous queries
  – Personalize: “uw”
  – Don’t personalize: “uw seattle library homepage”
• Identify easily personalized queries
  – Re-finding queries

Other Ways to Personalize

• Measuring the value of personalization
• Understanding the individual
• Calculating personal relevance
• Other ways to personalize search
  – Match expectation for re-finding queries
  – Personalized snippets

Ranking Results for Re-Finding
Interest-Based Snippets

Query: “winery”

Winery - Wikipedia, the free encyclopedia
A winery is a building or property that produces wine, or a business involved in the production of wine, such as a wine company. Some wine companies own many wineries. Besides wine making equipment ... en.wikipedia.org/wiki/Winery

If the person is interested in Maui:
Winery - Wikipedia, the free encyclopedia
A winery is a building or property that produces wine, or a business involved in the production of wine, such as a wine company... For example, in Maui there is a pineapple winery... en.wikipedia.org/wiki/Winery

Summary

• Measuring the value of personalization
  – There’s a big gap between group and individual
• Understanding the individual
  – Building a profile, explicit v. implicit
• Calculating personal relevance
  – Relevance feedback, boost click through
• Other ways to personalize search
  – Rank based on expectation, personalized snippets