Modeling User Behavior and Interactions

Lecture 4: Search Personalization

Eugene Agichtein
Emory University
1. Approaches to Search Personalization

2. Dimensions of Personalization
   1. Which queries to personalize?
   2. What input to use for personalization?
   3. Granularity: personalization vs. groupization
   4. Context: Geographical, search session
Approaches to Personalization

1. Pitkow et al., 2002
2. Qiu et al., 2006
3. Jeh et al., 2003
4. Teevan et al., 2005
5. Das et al., 2007

Figure adapted from: Personalized search on the world wide web, by Micarelli, A. and Gasparetti, F. and Sciarrone, F. and Gauch, S., LNCS 2007
When to Personalize

Figure adapted from: Personalized search on the world wide web, by Micarelli, A. and Gasparetti, F. and Sciarrone, F. and Gauch, S., LNCS 2007
Example: Outride

From Pitkow et al., 2002

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Outride (Results)

From Pitkow et al., 2002

<table>
<thead>
<tr>
<th>Search Engine</th>
<th>Average Task Completion Time in Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outride</td>
<td>38.9</td>
</tr>
<tr>
<td>Google</td>
<td>75.4</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>81</td>
</tr>
<tr>
<td>Excite</td>
<td>83.5</td>
</tr>
<tr>
<td>AOL</td>
<td>89.6</td>
</tr>
</tbody>
</table>

(% slower from Outride enabled search)

<table>
<thead>
<tr>
<th>Search Engine</th>
<th>User Actions</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outride</td>
<td>11.2</td>
<td>89.6</td>
</tr>
<tr>
<td>Google</td>
<td>21.2</td>
<td>100.5</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>22.4</td>
<td>108.5</td>
</tr>
<tr>
<td>AOL</td>
<td>23.1</td>
<td>107.0</td>
</tr>
<tr>
<td>Excite</td>
<td>23.3</td>
<td>114.5% slower</td>
</tr>
<tr>
<td>Average</td>
<td>22.5</td>
<td>101.4</td>
</tr>
</tbody>
</table>


(Average time to complete task)

From Pitkow et al., 2002
Input to Personalization

• Behavior (clicks): Qiu and Cho, 2006
  – Use clicks to tune a personalized (topic sensitive) PageRank model for each user
  – Use personalized PageRank to re-rank web search results

• Profile (user model): SeeSaw (Teevan et al., 2005)
PageRank Computation

I: Set of Incoming links
O: Set of Outgoing links
\( c \): Dampening factor (\( \sim 0.15 \)) or “teleportation probability”
E: Some probability vector over the Webpages

\[
PR(p) = (1 - c) \cdot \sum_{q \in I(p)} \frac{PR(q)}{|O(q)|} + c E(p)
\]

E vector can be:
- Uniformly distributed probabilities over all Web Page (democratic)
- Biased distributed probabilities to a number of important pages
  - Top-levels of Web Servers
  - Hub/Authority pages
- Used for Customization (Personalization)
Topic-Sensitive PageRank

Main Idea

- Assign *multiple a-priori “importance”* estimates to pages with respect to a set of *topics*
- One PageRank score per *basis topic*
  - Query specific rank score (+)
  - Make use of context (+)
  - Inexpensive at runtime (+)
PageRank vs Topic-Sensitive PageRank

**PageRank**

- **Input:** Web graph $G$
- **Output:** Rank vector $r : (\text{page} \rightarrow \text{page importance})$

**Topic-Sensitive PageRank**

- **Input:** Web $W$, Basis topics $[c_1, ... , c_{16}]$, e.g. 16 categories (first level of ODP)
- **Output:** List of rank vectors $[r_1, ... , r_{16}]$
  - $r_j : \text{page} \rightarrow \text{page importance in topic } c_j$

---

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Input to Personalization

• Behavior (clicks): Qiu and Cho, 2006
  – Use clicks to tune a personalized (topic sensitive) PageRank model for each user
    ➢ Map clicked results to ODP
  – Use personalized PageRank to re-rank web search results

• Profile (user model): SeeSaw (Teevan et al., 2005)
PS Search Engine (Profile-based)

[Teevan et al., 2005]

User profile: Content, interaction history
Result Re-Ranking

- Ensures privacy
- Good evaluation framework
- Can look at rich user profile
- Look at light weight user models
  - Collected on server side
  - Sent as query expansion

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
BM25 with Relevance Feedback

\[
\text{Score} = \sum tf_i \times w_i
\]

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

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
User Model as Relevance Feedback

Score = $\sum tf_i \times w_i$

$N' = N + R$

$n_i' = n_i + r_i$

$w_i = \log \frac{(r_i + 0.5)(N' - n_i' - R + r_i + 0.5)}{(n_i' - r_i + 0.5)(R - r_i + 0.5)}$

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
User Model as Relevance Feedback

**World Focused Matching**

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

**Query Focused Matching**

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
User Representation

- Stuff I’ve Seen (SIS) index
  - MSR research project [Dumais, et al.]
  - Index of everything a user’s seen
- Recently indexed documents
- Web documents in SIS index
- Query history
- None
World Representation

- Document Representation
  - Full text
  - Title and snippet
- Corpus Representation
  - Web
  - Result set – title and snippet
  - Result set – full text

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Parameters

- **Matching**
  - Query focused
  - World focused

- **User representation**
  - All SIS
  - Recent SIS
  - Web SIS
  - Query history
  - None

- **World representation**
  - Full text
  - Title and snippet

- **Query expansion**
  - Web
  - Result set – full text
  - Result set – title and snippet

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Results: *Seesaw* Improves Retrieval

![Bar Chart]

- **No user model**
- **Random**
- **Relevance Feedback**
- **Seesaw**

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Results: Feature Contribution

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Summary

- Rich user model important for search personalization
- Seesaw improves text based retrieval
- Need other features to improve Web
- Lots of room for improvement
Evaluating Personalized Search

• Explicit judgments (offline and in situ)
  – Evaluate components before system
  – NOTE: What’s relevant for you

• Deploy system
  – Verbatim feedback, Questionnaires, etc.
  – Measure behavioral interactions (e.g., click, reformulation, abandonment, etc.)
  – Click biases –order, presentation, etc.
  – Interleaving for unbiased clicks

• Link implicit and explicit (Curious Browser plugin)

• Beyond a single query -> sessions and beyond
User Control in Personalization (RF)

Study: Comparing Personalization Strategies

- 10,000 users, 56,000 queries, and 94,000 clicks over 12 days.
- Used the first 11 days' worth of data to form user profiles and clicks.
- Simulated the application of five different personalization algorithms on the remaining 4,600 queries from the last day of the log.
- Retrieved top 50 results for each query from the comparison search engine and assumed that clicking a link indicated a relevance judgment for the query.
Results: Which Strategy is Most Effective?

[ Dou et al., 2007]

- Compared two click-based (behavior) personalization strategies to three profile-based strategies
- Click-based strategies appear more effective than profile-based (but carefully combining historical profile data helps slightly)
- Search context crucial
- Personalization effectiveness varies by query

⚠ Evaluated using naïve click models
Approaches to Search Personalization

1. Dimensions of Personalization
   - What input to use for personalization?
   - Which queries to personalize?
   1. Granularity: personalization vs. groupization
   2. Context: Geographical
Understanding Query Ambiguity

SIGIR 2008

Jaime Teevan, Susan Dumais, Dan Liebling
Microsoft Research
“grand copthorne waterfront”
"singapore"
How Do the Two Queries Differ?

• grand copthorne waterfront v. singapore

• Knowing query ambiguity allow us to:
  – Personalize or diversify when appropriate
  – Suggest more specific queries
  – Help people understand diverse result sets

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Understanding Ambiguity

• Look at measures of query ambiguity
  – Explicit
  – Implicit

• Explore challenges with the measures
  – Do implicit predict explicit?
  – Other factors that impact observed variation?

• Build a model to predict ambiguity
  – Using just the query string, or also the result set
  – Using query history, or not
Which Queries to Personalize? [Teevan et al., 2008]

• Personalization benefits **ambiguous** queries

• Inter-rater reliability (Fleiss’ kappa)
  – Observed agreement \((P_a)\) exceeds expected \((P_e)\)
  – \(\kappa = \frac{(P_a-P_e)}{(1-P_e)}\)

• Relevance entropy
  – Variability in probability result is relevant \((P_r)\)
  – \(S = -\sum P_r \log P_r\)

• Potential for personalization
  – Ideal group ranking differs from ideal personal
  – \(P4P = 1 - nDCG_{\text{group}}\)

Teevan, J., S. T. Dumais, and D. J. Liebling. *To personalize or not to personalize: modeling queries with variation in user intent.*, SIGIR 2008
## Predicting Ambiguity

[Teevan et al., 2008]

<table>
<thead>
<tr>
<th>Information</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Query</strong></td>
<td>Query length, Contains URL, Contains advanced operator, Time of day issued, Number of results (df), Number of query suggests</td>
<td>Reformulation probability, # of times query issued, # of users who issued query, Avg. time of day issued, Avg. number of results, Avg. number of query suggests</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td>Query clarity, ODP category entropy, Number of ODP categories, Portion of non-HTML results, Portion of results from .com/.edu, Number of distinct domains</td>
<td>Result entropy, Avg. click position, Avg. seconds to click, Avg. clicks per user, <strong>Click entropy</strong></td>
</tr>
</tbody>
</table>

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Collecting Implicit Relevance Data

- Variation in clicks
  - Proxy (click = relevant, not clicked = irrelevant)
  - Other implicit measures possible
  - Disadvantage: Can mean lots of things, biased
  - Advantage: Real tasks, real situations, lots of data

- 44k unique queries issued by 1.5M users
  - Minimum 10 users/query

- 2.5 million result sets “evaluated”
How Good are Implicit Measures?

• Explicit data is expensive
• Implicit good substitute?
• Compared queries with
  – Explicit judgments and
  – Implicit judgments
• Significantly correlated:
  – Correlation coefficient $= 0.77 (p<.01)$
Which Has Lower Click Entropy?

- **www.usajobs.gov** v. federal government jobs
- find phone number v. **msn live search**
- **singapore pools** v. singaporepools.com

Click entropy = 1.5
Result entropy = 5.7

Click entropy = 2.0
Result entropy = 10.7

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Challenges with Using Click Data

• Results change at different rates
• Result quality varies
• Task affects the number of clicks

• We don’t know click data for unseen queries

➤ Can we predict query ambiguity?
Result Summary [Teevan et al., 2008]

- All features = good prediction
  - 81% accuracy (↑ 220%)
- Just query features promising
  - 40% accuracy (↑ 57%)
- No boost adding results or history

Teevan, J, S. T. Dumais, and D. J. Liebling. To personalize or not to personalize: modeling queries with variation in user intent., SIGIR 2008

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Lecture 4 Outline

✓ Approaches to Search Personalization

1. Dimensions of Personalization
   ✓ What input to use for personalization?
   ✓ Which queries to personalize?
   ➢ Granularity: personalization vs. groupization

1. Context: Geographical, search session
Connection: Collaborative Filtering and Recommender Systems

– Identify related groups

  • Browsed pages [Almeida & Almeida 2004; Sugiyama et al. 2005]
  • Queries [Freyne & Smyth 2006; Lee 2005]
  • Location [Mei & Church 2008], company [Smyth 2007], etc.

– Use group data to fill in missing personal data

  • Typically data based on user behavior
Discovering and Using Groups to Improve Personalized Search

Jaime Teevan, Merrie Morris, Steve Bush
Microsoft Research
WSDM 2009
Diego Velasquez, Las Lanzas

[ Slides from Teevan et al., WSDM 2009 ]
People Express Things Differently

• Differences can be a challenge for Web search
  – *Picture of a man handing over a key.*
  – *Oil painting of the surrender of Breda.*

• Personalization
  – Closes the gap using more about the person

• *Group*ization
  – Closes the gap using more about the *group*
How to Take Advantage of Groups?

- Who do we share interests with?
- Do we talk about things similarly?
- What algorithms should we use?
Approach

• Who do we share interests with?
  – Similarity in query selection
  – Similarity in what is considered relevant

• Do we talk about things similarly?
  – Similarity in user profile

• What algorithms should we use?
  – **Groupize** results using groups of user profiles
  – Evaluate using groups’ relevance judgments
Interested in Many Group Types

- Group longevity
  - Task-based
  - Trait-based
- Group identification
  - Explicit
  - Implicit
Queries Studied

Trait-based dataset
• Challenge
  – Overlapping queries
  – Natural motivation
• Queries picked from 12
  – Work
    - c# delegates, live meeting
  – Interests
    - bread recipes, toilet train
    - dog

Task-based dataset
• Common task
  – Telecommuting v. office
    - pros and cons of working in an office
    - social comparison
    - telecommuting versus office
    - telecommuting
    - working at home cost
    - benefit
Data Collected

- Queries evaluated
- Explicit relevance judgments
  - 20 - 40 results
  - Personal relevance
    - Highly relevant
    - Relevant
    - Not relevant
- User profile: Desktop index
Answering the Questions

- Who do we share interests with?
- Do we talk about things similarly?
- What algorithms should we use?
Who do we share interests with?

- Variation in query selection
  - Work groups selected similar work queries
  - Social groups selected similar social queries
- Variation in relevance judgments
  - Judgments varied greatly ($\kappa=0.08$)
  - Task-based groups most similar
  - Similar for one query $\neq$ similar for another
Do we talk about things similarly?

- Group profile similarity
  - Members more similar to each other than others
  - Most similar for aspects related to the group

<table>
<thead>
<tr>
<th></th>
<th>In task group</th>
<th>Not in group</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>All queries</td>
<td>0.42</td>
<td>0.31</td>
<td>34%</td>
</tr>
<tr>
<td>Group queries</td>
<td>0.77</td>
<td>0.35</td>
<td>120%</td>
</tr>
</tbody>
</table>

- Clustering profiles recreates groups
- Index similarity ≠ judgment similarity
  - Correlation coefficient of 0.09
What algorithms should we use?

- Calculate personalized score for each member
  - Content: User profile as relevance feedback
    \[ \sum_{terms} \frac{tf_i}{log} \frac{(r_i+0.5)(N-n_i-R+r_i+0.5)}{(n_i-r_i+0.5)(R-r_i+0.5)} \]
  - Behavior: Previously visited URLs and domains

- Sum personalized scores across group
- Produces same ranking for all members
Performance: Task-Based Groups

- Personalization improves on Web
- Groupization gains +5%

[ Slides from Teevan et al., WSDM 2009 ]
Performance: Task-Based Groups

- Personalization improves on Web
- Groupization gains +5%
- Split by query type
  - On-task v. off-task
  - Groupization the same as personalization for off-task queries
  - 11% improvement for on-task queries
Performance: Trait-Based Groups

Normalized DCG

Groupization
Personalization

[ Slides from Teevan et al., WSDM 2009 ]
Performance: Trait-Based Groups

[ Slides from Teevan et al., WSDM 2009 ]

- Interests
- Work

Normalized DCG

- Groupization
- Personalization

Work queries

Interest queries
Performance: Trait-Based Groups

[ Slides from Teevan et al., WSDM 2009 ]

Normalized DCG

Groupization

Personalization

Interests

Work

Work queries

Interest queries
Lecture 4 Outline

✓ Approaches to Search Personalization

1. Dimensions of Personalization

✓ What input to use for personalization?
✓ Which queries to personalize?
✓ Granularity: personalization vs. groupization

➢ Context: Geographical, search session
Local Search (Geographical Personalization)

• Location is context
• *Local search* uses geographic information to modify the ranking of search results
  – location derived from the query text
  – location of the device where the query originated
• e.g.,
  – “underworld 3 cape cod”
  – “underworld 3” from mobile device in Hyannis
Geography and Query Intent

[ Baeza-Yates and Jones] 2008

Location 1: query location
“Pizza Amherst, MA” query1

Distance 1: home–query intent

Location 2: Home address

IP address / profile zip

Distance 2: Reformulation distance

“Pizza Northampton” query2

Location 3: query location

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Topic-Distance Profiles
[Baeza-Yates and Jones] 2008

• 20 bins
  – 0 distance
  – Equal fractions of the rest of the data

• Does distribution into distance bins topics vary by topic?

Movie theater  Distant places  Near-by

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Lecture 4 Outline

✓ Approaches to Search Personalization

1. Dimensions of Personalization
   ✓ What input to use for personalization?
   ✓ Which queries to personalize?
   ✓ Granularity: personalization vs. groupization
   ✓ Context: Geographical, search session
Key References and Further Reading


<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Conference/Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teevan et al.</td>
<td>Personalizing search via automated analysis of interests and activities.</td>
<td>Proc. of SIGIR 2005</td>
</tr>
<tr>
<td>Dou et al.</td>
<td>A large-scale evaluation and analysis of personalized search strategies,</td>
<td>Proc. of WWW 2007</td>
</tr>
<tr>
<td>Teevan et al.</td>
<td>To personalize or not to personalize: modeling queries with variation in user intent.</td>
<td>Proc. of SIGIR 2008</td>
</tr>
<tr>
<td>Teevan et al.</td>
<td>Discovering and Using Groups to Improve Personalized Search.</td>
<td>WSDM 2009</td>
</tr>
</tbody>
</table>