Personalized Interactive Faceted Search

Jonathan Koren*, Yi Zhang*, and Xue Liu†

*University of California, Santa Cruz
†McGill University
Outline

• Introduce Faceted Search
• Identify Problems with Current FS Tech
• Propose a Solution
• Novel Evaluation Methodology
• Experiments
• Conclusions
Faceted Search is Everywhere
Formal Definition

• Interactive Structured Search Using Key-Value Metadata
• Parallel Hierarchies of Documents
• Point and Click Structured Query Generation
Problems

• Too Many Facets and Values

• Existing approach: Ad Hoc Value Presentation

• Proposed Solution: Personalization and Collaborative faceted search for interactive system utility optimization
Statistical Modeling Framework

• Document Model
• User Relevance Model
Document Model

- Docs are Unique Facet-Value Pairs
- Facets Come in Different Types
  - Facet-Type Suggests Statistical Model
- Docs Modeled as a Combination of Statistical Models
User Relevance Model

$$\theta_u = \{P(\text{rel} \mid u), P(x_k \mid \text{rel}, u), P(x_k \mid \text{non}, u)\}$$
User Collaboration

- $\Phi$ is the Conjugate Prior to $\theta_u$
- $\Phi$ Fills in Gaps in Individual User Models
Interface Evaluation

- User Studies are Expensive
- New Complementary Approach
- Expected User Interface Utility
- Simulated Interaction with Pseudousers
User Interface Utility

- Identify Types of Actions
- Assign Costs to Actions
- Reward for Relevant Docs Retrieved
- Calculate Utility for Entire Search Session
Expected User Interface Utility

\[ E[U] = \sum_{u \in U} \sum_{D \in D} E[U(u, D)] P(D | u) P(u) \]

\[ E[U(u, D)] = \sum_{t=0} \sum_{a \in A_t} R(q_{t+1}, a, q_t) P(q_{t+1} | a, q_t, u) \]
\[ P(a | q_t, u, D) P(q_t | q_{t-1}, u, D) \]
Assumptions

1. Users Need to Satisfy a Need with a Set of Documents
2. Users Can Recognize Relevant Documents and Facet-Value Pairs
3. Users Continue to Perform Actions Until Their Need is Met
Pseudousers

- Stochastic Users
- First-Match Users
- Myopic Users
- Optimal Users
### Stochastic Users

- Picks Relevant FVP at Random

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Nonrelevant (14 matches)</td>
</tr>
<tr>
<td>B</td>
<td>Relevant (17 matches)</td>
</tr>
<tr>
<td>C</td>
<td>Relevant (11 matches)</td>
</tr>
<tr>
<td>D</td>
<td>Nonrelevant (12 matches)</td>
</tr>
<tr>
<td>E</td>
<td>Nonrelevant (12 matches)</td>
</tr>
<tr>
<td>F</td>
<td>Relevant (15 matches)</td>
</tr>
<tr>
<td>G</td>
<td>Relevant (13 matches)</td>
</tr>
<tr>
<td>H</td>
<td>Nonelevant (4 matches)</td>
</tr>
<tr>
<td>I</td>
<td>Relevant (13 matches)</td>
</tr>
<tr>
<td>J</td>
<td>Nonrelevant (16 matches)</td>
</tr>
</tbody>
</table>
First-Match Users

- Scans list for Relevant FVPs from Top to Bottom, Picking the First

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Nonrelevant</td>
<td>(14 matches)</td>
</tr>
<tr>
<td>B</td>
<td>Relevant</td>
<td>(17 matches)</td>
</tr>
<tr>
<td>C</td>
<td>Relevant</td>
<td>(11 matches)</td>
</tr>
<tr>
<td>D</td>
<td>Nonrelevant</td>
<td>(12 matches)</td>
</tr>
<tr>
<td>E</td>
<td>Nonrelevant</td>
<td>(12 matches)</td>
</tr>
<tr>
<td>F</td>
<td>Relevant</td>
<td>(15 matches)</td>
</tr>
<tr>
<td>G</td>
<td>Relevant</td>
<td>(13 matches)</td>
</tr>
<tr>
<td>H</td>
<td>Nonelevant</td>
<td>(4 matches)</td>
</tr>
<tr>
<td>I</td>
<td>Relevant</td>
<td>(13 matches)</td>
</tr>
<tr>
<td>J</td>
<td>Nonrelevant</td>
<td>(16 matches)</td>
</tr>
</tbody>
</table>
Myopic Users

- Picks Relevant FVP that is Contained in the Least Number of Documents
Optimal Users

- Examines the Complete Interface
- Executes the Action that Maximizes the Utility

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Nonrelevant</td>
</tr>
<tr>
<td>B</td>
<td>Relevant</td>
</tr>
<tr>
<td>C</td>
<td>Relevant</td>
</tr>
<tr>
<td>D</td>
<td>Nonrelevant</td>
</tr>
<tr>
<td>E</td>
<td>Nonrelevant</td>
</tr>
<tr>
<td>F</td>
<td>Relevant</td>
</tr>
<tr>
<td>G</td>
<td>Relevant</td>
</tr>
<tr>
<td>H</td>
<td>Nonelevant</td>
</tr>
<tr>
<td>I</td>
<td>Relevant</td>
</tr>
<tr>
<td>J</td>
<td>Nonrelevant</td>
</tr>
</tbody>
</table>
Evaluation Review

• Each Pseudouser Logs into the Search Interface

• Pseudouser Interacts with Interface to Retrieve a Set of Documents.

• Interface Receives a Score for the Session.

• Expected Utility = Average Score for all Sessions
Personalization Experiments

- Facet-Value Pair Suggestion
  - Most Frequent
  - Most Probable (Collaborative)
  - Most Probable (Personalized)
  - Mutual Information

- Start Page Personalization
  - Empty Page
  - Collaborative Page
  - Personalized page
Document Corpora

- 8000 Documents from IMDB
- 19 Facets and 367k Facet-Value Pairs
- 5000 Users Each from Netflix and MovieLens
  - 633k Ratings for Netflix
  - 742k Ratings for Movielens
Results (Netflix)

Averaged Num Actions vs. FVP Suggestion Method

- First-Match (Null Start)
- First-Match (Collab Start)
- First-Match (Personal Start)
- Myopic (Null Start)
- Myopic (Collab Start)
- Myopic (Personal Start)

Frequency, Collab Prob, Personal Prob, PMI
Results
(MovieLens)

- First-Match (Null Start)
- First-Match (Collab Start)
- First-Match (Personal Start)
- Myopic (Null Start)
- Myopic (Collab Start)
- Myopic (Personal Start)

Ave Num Actions

FVP Suggestion Method

Frequency Collab Prob Personal Prob PMI
Conclusions

• Many Facets and Values are a Problem
• Personalized Interfaces Can Help
• Proposed Statistical Modeling Framework for Faceted-Search
• Proposed Inexpensive Repeatable Evaluation Technique for Faceted-Search Interfaces
• Personalized Start Pages are Helpful
fin
Example: Two Myopic Users

Search for “The ‘Burbs”

User: 302

certificate=PG
soundmix=Dolby
genre=Comedy

User: 1329

certificate=PG
soundmix=Dolby
genre=Comedy
country=USA
language=English
colorinfo=Color
year=1989
productiondesigner=SpencerJamesH