Evaluating Preference-based Search Tools: A Tale of Two Approaches

Paolo Viappiani,¹ Boi Faltings,¹ and Pearl Pu²
¹Artificial Intelligence Lab (LIA)
²Human Computer Interaction Group (HCI)
Ecole Polytechnique Federale Lausanne (EPFL)

AAAI 2006, The Twenty-First National Conference on Artificial Intelligence, July, 16-20, Boston, MA, USA
Preference-based Search

- People often use the **WWW** to search for their most preferred item
  - Computers, cameras, apartments, flights
  - Structured items can be searched in a database

- **Crucial:** an accurate model of users’ preferences
  - Classic procedures for utility elicitation (Keeney) require too much effort
  - Most common approach is to ask the user to fill in a form
Example

- Actual scenario with travel website (July 5th, 2006)
- User wants to travel from Geneva to Dublin
- Return flight
- Preferences
  - Outbound flight, arrive by 5pm
  - Inbound flight, arrive by 3pm
  - (Cheapest)
Swiss will be cheaper
To be there at 5pm, I should leave around noon.
To arrive back at 3pm, I should leave in the morning.

<table>
<thead>
<tr>
<th>Depart</th>
<th>Arrival</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dublin [DUB] 12 Jul 06:45</td>
<td>Geneva [GVA] 12 Jul 12:45</td>
<td>05h 00m / 1 Stop / via Frankfurt Main [FRA] Economy</td>
</tr>
</tbody>
</table>

Fare per person: 2351 CHF (excl. taxes and fees)
Total for all passengers: 2610 CHF (incl. taxes and fees)
<table>
<thead>
<tr>
<th>Depart</th>
<th>Arrival</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dublin [DUB] 12 Jul 08:50</td>
<td>Geneva [GVA] 12 Jul 16:30</td>
<td>06h 40m / 1 Stop / via London [LHR] Economy</td>
</tr>
</tbody>
</table>

Fare per person: 1030 CHF (excl. taxes and fees)
Total for all passengers: 1217 CHF (incl. taxes and fees)

Change dates
- Leaving from: Geneva
- Going to: Dublin [DUB]
- Outbound: Monday hours 10/07/06 1200
- Inbound: Wednesday hours 12/07/06 0900
- Direct flights only
- 1 Adult(s)
- 0 Children
- 0 Infant(s)
- Preferred Airlines: All

Leave out preference about SWISS. Still expensive but cheaper; does not arrive at the preferred time.
Omit preference about departure time.

Much cheaper.

Outbound arrives by 5pm; Return arrives by 3pm, as desired.
Form-filling is not effective

- Incorrect means objectives: formulate the real goal by a “substitute” goal believed to lead to desired outcome

- Users often state more preferences than necessary when prompted

- *The preference model may be complete, but not accurate*
Alternative: preference construction

- Users’ preferences are often constructed when considering specific examples
  - behavioral decision theory (Payne et al. ’93; Slovic’95; Tversky ’96)

- Collaborative filtering recommends items based on users’ rating on similar items
  - When users volunteer to rate items, more accurate recommendations are given (McNee et al. ’03)

Allow users to self-initiate preference expression
**Example-based tools**

- Several proposed systems:
  - Findme (*Burke et al. ‘97*)
  - Smartclient (*PuFaltings’00*)
  - Expertclerk (*Shimazu’01*)

- User expresses the preferences as critiques on displayed examples
- Feedback directs the next search cycle
- Users are more motivated to express preferences when self-initiated
  - *Suggestions*

---

**Diagram:**
- **Initial preference**
- **The system shows K solutions**
- **The user critiques the solutions by stating a new preference**
- **The user picks the final choice**
The need for Suggestions

- Others have also recognized the need to help users consider potentially neglected attributes
- Show extreme examples (Linden’97)
- Show diverse examples (Smyth & McGinty’03, McSherry’02)
- Show suggestions based on the current preference model and possible extensions (Pu et al. ’06): model-based suggestion
  - Optimally stimulate preference expression
  - Metaphor of Active Learning
Example Critiquing vs. Form Filling

Via user studies, we ask

- Do EC tools achieve better decision accuracy than traditional form-filling approaches?
- Are preferences more accurate when they were obtained from example critiquing vs. form-filling?
User Studies

- **60 users** searched their most preferred item from **180 items** in a database.

- Measured variables
  - **Decision accuracy** (Pu&Chen ’05): the percentage of times the system succeeded in finding users’ most preferred item.
  - **User effort**: the task time a user takes while using the tool to reach an option that she believes to be the target item.
User Studies: Experiments

- **Between-groups experiment (3 groups of 20 people)**
  - **Form-filling interface**: user selects a preferred value or “don’t care” choice on each attribute
  - **Example-critiquing interface**: user only states self-initiated preferences; views 6 best options
  - **Example-critiquing interface with suggestions**: user only states self-initiated preferences; views 3 best options and 3 suggestions

- **Within-subject experiment (20 users)**
  - **Form-filling interface**
  - **Example-critiquing interface**: showing 3 best options and 3 suggestions
Flat Finder - Example Critiquing

Preferences

Price? 700

Distance to University? 10

Search according to these preferences: Search

Add preferences: Type Add

Results

In the dataset are present 21 accommodations opportunities. Of which 9 fully match your preferences.

These are the best solutions that match your query.

<table>
<thead>
<tr>
<th>ID</th>
<th>Type</th>
<th>Price</th>
<th>Rooms</th>
<th>Furnished</th>
<th>Smoking</th>
<th>Kitchen</th>
<th>Transportation</th>
<th>Distance to University</th>
<th>Choose</th>
<th>Distance to Centre</th>
</tr>
</thead>
<tbody>
<tr>
<td>7996</td>
<td>shared apartment</td>
<td>575</td>
<td>1.0</td>
<td>true</td>
<td>shared</td>
<td>shared</td>
<td>bus</td>
<td>9</td>
<td>Remove</td>
<td>8</td>
</tr>
<tr>
<td>8083</td>
<td>shared apartment</td>
<td>475</td>
<td>1.0</td>
<td>true</td>
<td>shared</td>
<td>shared</td>
<td>bus</td>
<td>10</td>
<td>Remove</td>
<td>12</td>
</tr>
<tr>
<td>7999</td>
<td>room in a house</td>
<td>490</td>
<td>1.0</td>
<td>false</td>
<td>either</td>
<td>shared</td>
<td>bus</td>
<td>10</td>
<td>Remove</td>
<td>12</td>
</tr>
</tbody>
</table>

In the dataset you can also find:

<table>
<thead>
<tr>
<th>ID</th>
<th>Type</th>
<th>Price</th>
<th>Rooms</th>
<th>Furnished</th>
<th>Smoking</th>
<th>Kitchen</th>
<th>Transportation</th>
<th>Distance to University</th>
<th>Choose</th>
<th>Distance to Centre</th>
</tr>
</thead>
<tbody>
<tr>
<td>3084</td>
<td>studio</td>
<td>550</td>
<td>1.0</td>
<td>false</td>
<td>either</td>
<td>private</td>
<td>private bus</td>
<td>14</td>
<td>Remove</td>
<td>3</td>
</tr>
<tr>
<td>7994</td>
<td>room in a house</td>
<td>550</td>
<td>1.0</td>
<td>true</td>
<td>shared</td>
<td>none</td>
<td>private bus</td>
<td>7</td>
<td>Remove</td>
<td>5</td>
</tr>
<tr>
<td>7992</td>
<td>apartment</td>
<td>625</td>
<td>1.5</td>
<td>false</td>
<td>either</td>
<td>private</td>
<td>private metro</td>
<td>9</td>
<td>Remove</td>
<td>8</td>
</tr>
</tbody>
</table>

Look at the solutions displayed. If you realize that you did not state some of your preferences you can do it now.

State an additional criterion: Type Add

My Basket

Here you can store entries for comparison. When you choose one of them, you can proceed to checkout.

No element in that set
Between groups Experiment

Accuracy increases with suggestions

- EC with suggestion
  - Better than form-filling (p<0.01)
  - Better than EC without suggestions (p<0.02)

- EC without suggestions
  - Better than Form filling (but p>0.05)

![Bar chart showing accuracy of different methods](chart.png)

<table>
<thead>
<tr>
<th></th>
<th>Form filling</th>
<th>Form &amp; revisions</th>
<th>Example-critiquing</th>
<th>EC + suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2:45</td>
<td>5:30</td>
<td>8:09</td>
<td>7:39</td>
</tr>
<tr>
<td>Cycles</td>
<td>1.0</td>
<td>2.2</td>
<td>5.6</td>
<td>6.3</td>
</tr>
</tbody>
</table>
Within-subject Experiment

EC with suggestions achieved better decision accuracy than simple form-filling (p<0.01) and repeated form-filling (p<0.03).

Accuracy is not due to more interaction.

<table>
<thead>
<tr>
<th></th>
<th>Time (avg)</th>
<th>Cycles (avg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form filling</td>
<td>2:45</td>
<td>1.0</td>
</tr>
<tr>
<td>Form &amp; revisions</td>
<td>5:30</td>
<td>2.2</td>
</tr>
<tr>
<td>EC + suggestions</td>
<td>6:00</td>
<td>5.2</td>
</tr>
</tbody>
</table>
Comments on the results

- **Form-filling**: users state average of 7.5 preferences
  - Before having considered any of the available options
  - Even after revisions, preferences were not retracted

- **EC**: users begin with average of only 2.7 preferences, added average of 2.6 to reach 5.3
  - 50% preferences were added during interaction
  - Results suggest that volunteered preferences are more accurate

- More preference revisions → higher decision accuracy (Pu&Chen ’05)
  - People who found their targets made more revisions
  - 6.9 as opposed to 4.5, statistically significant (p=0.0439)
Conclusions: a tale of two approaches

- Do not ask too many questions
  - Even though form filling interfaces are easier to implement

- Show attractive suggestions
  - User effort should be compatible with motivation for decision accuracy
  - Model-based suggestions effectively stimulate users to express accurate preferences

- User study validates the hypotheses
• Initial preference: lowest price

Other (hidden) preferences:
• Arrive by 12:00
• Leave from City airport

<table>
<thead>
<tr>
<th></th>
<th>Fare ($a_1$)</th>
<th>Arrival ($a_2$)</th>
<th>Airport ($a_3$)</th>
<th>Airline ($a_4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>250</td>
<td>14:00</td>
<td>INT</td>
<td>B</td>
</tr>
<tr>
<td>O2</td>
<td>300</td>
<td>9:00</td>
<td>INT</td>
<td>A</td>
</tr>
<tr>
<td>O3</td>
<td>350</td>
<td>17:30</td>
<td>CITY</td>
<td>B</td>
</tr>
<tr>
<td>O4</td>
<td>400</td>
<td>12:30</td>
<td>CITY</td>
<td>B</td>
</tr>
<tr>
<td>O5</td>
<td>550</td>
<td>18:30</td>
<td>CITY</td>
<td>B</td>
</tr>
<tr>
<td>O6</td>
<td>600</td>
<td>8:30</td>
<td>CITY</td>
<td>A</td>
</tr>
</tbody>
</table>

- Arrive by 12:00
- Leave from City airport
<table>
<thead>
<tr>
<th></th>
<th>Fare (a1)</th>
<th>Arrival (a2)</th>
<th>( \delta_2 )</th>
<th>Airport (a3)</th>
<th>( \delta_3 )</th>
<th>Airline (a4)</th>
<th>( \delta_4 )</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>250</td>
<td>14:00</td>
<td>-</td>
<td>INT</td>
<td>-</td>
<td>B</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>O2</td>
<td>300</td>
<td>9:00</td>
<td>0.5</td>
<td>INT</td>
<td>0</td>
<td>A</td>
<td>0.5</td>
<td>0.437</td>
</tr>
<tr>
<td>O3</td>
<td>350</td>
<td>17:30</td>
<td>0.35</td>
<td>CITY</td>
<td>0.5</td>
<td>B</td>
<td>0</td>
<td>0.381</td>
</tr>
<tr>
<td>O4</td>
<td>400</td>
<td>12:30</td>
<td>0</td>
<td>CITY</td>
<td>0</td>
<td>B</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>O5</td>
<td>550</td>
<td>18:30</td>
<td>0.1</td>
<td>CITY</td>
<td>0</td>
<td>B</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>O6</td>
<td>600</td>
<td>8:30</td>
<td>0.05</td>
<td>CITY</td>
<td>0</td>
<td>A</td>
<td>0</td>
<td>0.025</td>
</tr>
</tbody>
</table>

\[
P(o) = 1 \prod_{a_i \in A_u} (1 - P_{a_i \delta_i}(\theta))
\]