The Potential for Personalizing Search Support According to Interactive Search Intentions

Nicholas J. Belkin
Department of Library and Information Science
School of Communication & Information
Rutgers University, New Brunswick, NJ USA
belkin@rutgers.edu
How Personalization of Search Support is Typically Construed

• Tailoring search results in response to a query according to:
  • Physical and environmental conditions (e.g. geographic location)
  • Temporal conditions (e.g. time of day)
  • Past behaviors and, to some extent, current behaviors (what’s been liked before, what’s been open for more than some time period) of person, and persons “like” the person

• Interpreting and modifying a query according to:
  • Factors listed above;
  • Inferring original query “intent” (i.e. topic of search) on linguistic and behavioral evidence
  • Inferring search “intent” (type of search; goal of search) on linguistic and behavioral evidence
General Characteristics of this Type of Personalization

• Typically assumes a “static” person and goal
• Typically works within the context of a single query-response cycle
• Success evaluated according to document relevance
• Personalization of these sorts has been shown to be differentially useful, according to type of search/goal intent (Teevan, Dumais & Horvitz (2010) Potential for personalization, ACM TOIS).
Another Type of Personalization

• People engage in information search and seeking *sessions*,
• as a consequence of recognition that they are in a *problematic situation*,
• with respect to some *task* or *goal*
• People evaluate seeking and search support according to *usefulness* of sessions with respect to motivating task/goal
• People *change* during the course of seeking and searching sessions
• People try to realize different *intentions* during the course of information seeking and searching intentions, in furtherance of the (perhaps dynamic) goal of the session
• Personalization could be with respect to motivating goal/task, *and also* with respect to search intentions
• **Personalization with respect to support for interaction**
A Model of IR as Support for Interaction (Belkin, 1996)
A Program in Personalization for Interaction

• A research program at Rutgers University Department of Library and Information Science
• Personalization of the Digital Library Experience (POoDLE) - IMLS
• Automatic Identification of Information Searcher Intentions During an Information Seeking Session – Google
• Characterizing and Evaluating Whole Session Interactive Information Retrieval (CHEWS-IIR) – NSF (in progress, our focus today).
PoODLE Goals

• Identification, from information search behaviors, of
  • Motivating task type
  • Searcher domain knowledge
  • Difficulty in searching
• Investigating the influence of individual characteristics on search behaviors
  • Cognitive abilities
  • Knowledge
• Predicting document usefulness on the basis of search behaviors
Google Project Goals

• Identifying relationships of behaviors to search intentions
  • Especially, eye-fixation behaviors

• Classifying sequences of search behaviors

• Investigating differences in search intention-related behaviors between motivating task types
General Pattern of our Studies

• Construct work tasks of different types, with associated information searching tasks
• Have participants conduct search for one work task
  • Log behaviors
  • Record search session
• Play back information search session for participant annotation
• Iterate for next work task, to final work task
• Exit interview
Work Tasks and Information Search Tasks

• Journalism Domain
  • Any topic
  • Several well-defined types of work tasks, e.g.
    • Advance obituary; Copy editing; Prepare for interview; Story pitch; Prepare story

• Constructed work and search tasks differ on values of specific facets
  • Faceted classification of task (Li & Belkin, 2008)
Li & Belkin (2008) Facet Analysis of Task (modified)

- Source of Task
  - Self, Group, Assigned
- Task Doer
  - Individual, Group
- Time
  - Frequency
  - Length
  - Stage
- Product
  - Physical, Intellectual, Decision, Factual
- Process
  - One-time, Multiple

- Items
  - Named or Not
  - Whole or Part
- Goal
  - Quality
    - Specific, Amorphous, Mixed
  - Quantity
    - Single or multiple goals
- Common attributes of task, e.g.
  - Objective/Subjective task complexity, Urgency, Salience, Difficulty, ...
Assignment 1. Copy Editing (CPE)
Your Assignment: You are a copy editor at a newspaper and you have only 20 minutes to check the accuracy of six italicized statements in the excerpt of a piece of news story below.
Your Task: Please find and save an authoritative page that either confirms or disconfirms each statement.
Product: Fact; Items: Named/Part; Goal: Specific
Example Task and Classification

Assignment 2. Story Pitch (STP)
Your Assignment: You are planning to pitch a science story to your editor and need to identify interesting facts about the coelacanth (“see-la-kanth”), a fish that dates from the time of dinosaurs and was thought to be extinct.
Your Task: Find and save web pages that contain the six most interesting facts about coelacanths and/or research about their preservation.
Product: Fact; Items: Not Named/Part; Goal: Specific
Assignment 3. Relationships (REL)

Your Assignment: You are writing an article about coelacanths and conservation efforts. You have found an interesting article about coelacanths but in order to develop your article you need to be able to explain the relationship between key facts you have learned.

Your Task: In the following there are five italicized passages, find an authoritative web page that explains the relationship between two of the italicized facts.

Product: Intellectual; Items: Named/Part; Goal: Mixed (Specific + Amorphous)
Assignment 4. Interview Preparation (INT)

Your Assignment: You are writing an article that profiles a scientist and their research work. You are preparing to interview Mark Erdmann, a marine biologist, about coelacanths and conservation programs.

Your Task: Identify and save authoritative web pages for the following:
- Identify two (living) people who likely can provide some personal stories about Dr. Erdmann and his work.
- Find the three most interesting facts about Dr. Erdmann’s research.
- Find an interesting potential impact of Dr. Erdmann’s work.

Product: Intellectual; Items: Not-Named/Whole; Goal: Amorphous
Participants and Procedure

- Journalism undergraduate university students (N=40)
- Entry questionnaire – demographics
- Searches for two (of four) tasks conducted in lab with eyetracker (20 minutes each)
- Pre-search questionnaire (when presented with task description)
  - Familiarity with task, topic
  - Expected difficulty
- Search conducted on Web, any search system, through Coagmento
- Post-search questionnaire
  - Experienced difficulty
  - Confidence in task success
- Playback search for annotation, by Query Segment
  - QS is query n, all that happens up to and including query n+1 (or end)
- Exit interview
  - Comparison of two tasks and two search sessions
Annotation

• Play back QS n
• What were you intending to accomplish during this period
  • Choice of intentions, can be multiple
• For each intention: Was this intention satisfied? If no, why not
  • [text entry]
• What were you hoping to accomplish with [query n+1]
  • [text entry]
• Play back QS n+1
Xie’s (2002) Interactive [Search] Intentions

• Identify search information (Something to start; Something more to search)
• Learn (Domain knowledge; Database content)
• Find (Known item; Specific information; Sharing named characteristic; Without predefined criteria)
• Keep record
• Access item or set of items
• Evaluate (Correctness; Usefulness; Best; Specificity; Duplication)
• Obtain (Specific information; Part of item; Whole item)
Data Analyses (So Far)

• Querying behavior and search intentions
  • Relationships between query reformulation “types” and search intentions
  • Effect of intention satisfaction on query reformulation type
  • Classification of reasons for query reformulation

• Intentions and search behaviors (again, our focus today)
  • Are the Xie search intentions necessary and sufficient
  • Sequences of search intentions
  • Prediction of search intention based on search behavior
Are Xie’s (2002) Intentions Necessary and Sufficient?
Necessary: Yes; Sufficient: Maybe

• Despite the nature of the work and search tasks, participants had no difficulty identifying different intentions associated with different query segments.

• Given the different nature of the various intentions, this suggests that search support techniques other than query reformulation could be useful in supporting effective interaction.

• The degree of satisfaction of intentions may be due to either low expectations, or inventive use of reformulation.
Search Intentions and Search Behaviors

Given that people attempt to accomplish different intentions during the course of an information search session, can a system identify what those intentions are, without intervention?

• RQ1: How is a user’s Web search behavior associated with his or her information seeking intentions in the same query segment
• RQ2: How is a user’s Web search behavior in the current query segment associated with his or her information seeking intentions in the subsequent query segment
Observed Behaviors, per Query Segment

- Saved item (binary)
- Number of saved items
- Dwell times on content pages
- Dwell times on SERP viewports
- Query length
- Query reformulation type
- Number of clicks
- Number of sources visited
- Number of pages viewed

- Dwell times are:
  - total dwell time
  - total dwell time until a page is saved
  - total open time
  - total open time until a page is saved
  - first dwell time
  - mean of all dwell times
Behavioral Groups for Binary Classification
(Tested singly and in combinations)

• Saving features
  • Saved item (binary)
  • Number of saved items

• Content page features
  • Dwell times
  • Number of content pages, by types: saved, not saved, unsaved, total

• SERP (i.e. viewport on SERP) features
  • Dwell times

• Query features
  • Query length
  • Query reformulation type
Measuring Performance

• Measures
  TP=True Positive; FP=False Positive;  
  TN=True Negative; FN=False Negative

• Accuracy:
  \[ \text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \]

• Precision for intention present:
  \[ P_1 = \frac{TP}{TP + FP} \]

• Precision for intention absent:
  \[ P_0 = \frac{TN}{TN + FN} \]

• Baselines
  • Stratified sampling of positive / negative labels proportional to their distribution in training data
  • Assigning the most frequent label in the training data

• Tests for Identification
  • Improvement over the better of the two baselines, Kolmogorov-Smirnoff
Results for Identification by Classification (1)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Range [min, max]</th>
<th>Range Increase [min, max]</th>
<th>Mean</th>
<th>Mean Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>[.608, .929]</td>
<td>[0, .129]</td>
<td>.786</td>
<td>.024</td>
</tr>
<tr>
<td>$P_0$</td>
<td>[.661, .967]</td>
<td>[0, .545]</td>
<td>.808</td>
<td>.064</td>
</tr>
<tr>
<td>$P_1$</td>
<td>[.411, 1.0]</td>
<td>[.044, .91]</td>
<td>.768</td>
<td>.522</td>
</tr>
</tbody>
</table>

- For all intentions, Accuracy and $P_0$ significantly improved over better baseline (mostly $p < .01$), when using all features together.
Results for Identification by Classification (2)

• Accuracy
  • Significant (p < .01) but not large improvement in ACC over better baseline for all intentions but one. For most intentions, using all feature groups was best

• Precision present
  • Significant (p < .01) and meaningful improvement in $P_{\text{pres}}$ for all intentions; For most intentions, one, or a combination of two feature groups performed best, rather than combining all.

• Precision absent
  • Slight improvements, most non-significant, over best baseline. Scores were uniformly fairly high for both baseline
Classification Discussion

• Doing better than random with a very simple classifier for two out of three measures

• Doing very well in Positive identification, likely because it’s a conservative algorithm
  • Identifying fewer intentions, with more certainty, is probably a win given the problem

• Negative identification may be uninteresting, given the problem

• Interesting start on the problem; next steps are:
  • More and different features
  • Prediction, rather than just identification
Behavioral Groups for Prediction

- Overall search behavior
  - Query length
  - Number of sources visited
  - Number of pages viewed

- Dwell time features
  - Mean dwell time on each SERP viewport
  - Mean dwell time on content pages

- Usefulness judgment
  - Saved item (binary)
  - Number of saved items
Measuring Strength of Relationship

• RQ 1: How is a user’s Web search behavior associated with his or her information seeking intentions in the same query segment
  • Mean value of each search behavior for all query segments
  • Mean value of each search behavior for query segment with given intention
  • Degree of difference between the two indicates strength of relationship

• RQ2: How is a user’s Web search behavior in the current query segment associated with his or her information seeking intentions in the subsequent query segment
  • Mean value of each search behavior for all query segments
  • Mean value of each search behavior for query segment preceding query segment with given intention
  • Degree of difference between the two indicates strength of relationship
Methods

- Correlation analysis for each behavior-intention pair
- Done for all current, and all subsequent, pairs
- Behaviors distributed non-normally
- Mann-Whitney tests for significant differences
Results for Identification and Prediction by Deviation from Mean

• In general, different behaviors, and patterns of behaviors, are associated with different intentions in the current query segment. Many significant such associations

• In general, different behaviors, and patterns of behaviors, in the current query segment are associated with different intentions in the subsequent query segment. Fewer significant such associations than for current intention, but still some for almost all subsequent intentions

• Next two slides show these results for (1) identification and (2) prediction. Black is significantly above the mean; grey is significantly below the mean
<table>
<thead>
<tr>
<th>Current intention</th>
<th>Query</th>
<th>Clicks</th>
<th>Sources</th>
<th>Pages</th>
<th>SERP</th>
<th>Content</th>
<th>if bookmark</th>
<th># bookmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>website</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>common</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>item</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>best</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>correctness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>duplicate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>usefulness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find</td>
<td>common</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>known</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>without</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify</td>
<td>more</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>start</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keep</td>
<td>link</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learn</td>
<td>database</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtain</td>
<td>part</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>whole</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Next intention</td>
<td>Query</td>
<td>Clicks</td>
<td>Sources</td>
<td>Pages</td>
<td>SERP</td>
<td>Content</td>
<td>if bookmarked</td>
<td># bookmarks</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>--------</td>
<td>---------</td>
<td>-------</td>
<td>------</td>
<td>---------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Access</td>
<td>website</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>common</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>item</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>best</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>correctness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>duplicate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>usefulness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Find</td>
<td>common</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>known</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>without</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify</td>
<td>more</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>start</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keep</td>
<td>link</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learn</td>
<td>database</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtain</td>
<td>part</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>whole</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Next Steps

• Adding analysis of eye fixation behaviors to the identification and prediction models

• Carry out analyses with respect to task types and facet values
  • Substantial evidence that task type influences search behaviors significantly
  • Strong suspicion that task type influences patterns of intentions

• Carrying out *in situ* study of search behaviors and search intentions
  • Thirty “professional” participants, searches logged and annotated by intentions, for one week.
Thanks for Your Attention

• Acknowledgements due to all of the members of the POoDLE and CHEWS-IIR projects, and to our funders

• Work reported here was supported through the National Science Foundation, grant #IIS-1423239.

• Work reported here was supported by a Google Faculty Research Award to N.J. Belkin & C. Shah

• Some work reported here was supported by IMLS grant LG#06-07-0105-07