Scholarly Usage Based Recommendations: Evaluating bX for a Consortium

Dana Thomas, Scholars Portal
Amy Greenberg, Scholars Portal
Pascal Calarco, University of Waterloo

Igelu, Haifa, September 2011
Agenda

• Recommender systems and evaluation methods overview
• Introduction to SFX, Primo, and bx in OCUL
• Findings to date in OCUL usage data
• Focus groups at Waterloo
• Future plans for bx evaluation at OCUL
| Scholarly Recommender Systems | Content-based | - STM digital collections  
- Considers bibliographic citations in articles as ‘preferences’ | - Tested 1886 articles which generated 1.9 million recommendations  
- Compared to bX:  
- Produced similar # recommendations  
- High ‘semantic diversity’ of recommendations |
|-----------------------------|---------------|-----------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Synthese (Vellino, 2010)    | Content-based | - Groups of researchers and resources (categorised via ScienceDirect)  
- If algorithm determines that user-suggested new resources relevant for wave/s, suggestion sent to wave administrator for approval | - Prototype tested 7 waves of 34 researchers and 83 resources  
- Resources considered “well-inserted” by wave administrators, with low percentage of missed resources |
| Google Wave-based system    | Content-based | - Uses overall context of document and local surrounding words to produce ranked set of citations | - Tested 450,000 pre-2008 articles  
- Title and abstract used for global context  
- 50 words before and after placeholder used as local citation context  
- Evaluated 3 measures of relevance and found to be effective compared to other systems |
| Serrano-Guerrero et al., 2011 | Content-based | - MeSH headings mapped to 130 resources and user queries  
- Recommendations generated based on frequency of common headings returned | - Goal was promotion of library resources  
- Measured user click-through rates to suggested resources  
- Statistically inconclusive positive trend |
## Scholarly Recommender Systems (cont’d)

<table>
<thead>
<tr>
<th>System</th>
<th>Method</th>
<th>Usage Details</th>
<th>Testing Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bx (Bollen &amp; Van de Sompel, 2006)</td>
<td>Usage-based</td>
<td>- Usage data for user community aggregated by openURL linking</td>
<td>- Tested using SFX linkage data from 9 CalState institutions and 2 million unique referents</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Data harvested by OAI-PMH</td>
<td>- Journal relationships established via PageRank algorithm</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Potential privacy issue can be resolved using anonymized session IDs</td>
</tr>
<tr>
<td>Melvyl Recommender Project (Whitney &amp; Schiff, 2006)</td>
<td>Usage-based</td>
<td>- Anonymous but persistent patron IDs</td>
<td>- Checkout = +ve rating??</td>
</tr>
<tr>
<td></td>
<td>(circulation data from UCLA)</td>
<td>- Large volume of data (~9 million transaction records)</td>
<td>- Small user study (10 undergrads and grad students)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Weighted recommendations based on common checkouts, filtered by call number similarity</td>
<td>- Survey and think-aloud tasks</td>
</tr>
<tr>
<td>Personal Ontology Recommender (PORE) system (Liao et al., 2010)</td>
<td>Hybrid</td>
<td>- Uses loan history of patron to establish “personal ontology” of CCL or DDC categories and keywords used</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Finds similar users based on keyword overlap in loan records</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Recommends items based on frequency of keyword overlap</td>
<td></td>
</tr>
</tbody>
</table>
Evaluation Approaches for Recommender Systems

• One formula to rule them all? (del Olmo & Gaudioso, 2008)
• Offline and online tests
• User studies
• Properties of interest should be identified first (Shani & Gunawardana, 2011):
  – Accuracy, ranking, novelty, serendipity, coverage, confidence, trust, diversity, utility, risk, robustness, privacy, adaptivity, scalability
Previous Evaluations

• Bx:
  – Kansas State, presented at ELUNA, faculty survey delivered through email
  – CISTI comparison of Synthese (their content-based system) and bx

• Melvyl Project:
  – CDL OPAC recommender system evaluated through focus group. Tested both collaborative filtering and metadata/content-based recommenders.
Our Context: SFX bx and Primo in OCUL

• OCUL is a consortium of 21 university libraries in Ontario with a shared technology infrastructure
• Licensed and locally hosted SFX instances for all members since 2004
• Current Primo beta at the TUG group (Guelph, Waterloo and Wilfrid Laurier)
• Various discovery systems at other schools
• TUG group started bx trial in summer 2010  The rest of OCUL began February 2011
• Activated by 8/21 OCUL schools to date.
Configuration Options used in OCUL

- Display SFX button for all recommendations: 8/8
- Recommend only full text articles: 1/8 (Waterloo)
- Display “full text available” image: 2/8
- Direct link to full text articles: 2/8
- Contributing Data: 0/8
- Most show default 3 recommendation, Carleton shows 10.
Approach to Evaluating bx in OCUL

- SFX queries: 1, 2, 4, 11
  - Frequency of recommendations generated
  - Frequency recommendations clicked on
  - Ratio of recommendations followed: menus displaying recommendations
  - Titles found through bx
  - Relationship (?) between bx and ILL requesting
  - Relationship (?) between recommendations and other SFX target service use
Approach to Evaluation bx in OCUL

– Relationship (?) between configuration choices and get recommendation clickthroughs

• Focus Groups (include undergrads next)

• Surveys
  – Building on work at Kansas State by Jamene Brooks-Kieffer
OCUL Usage Data Findings

• 6 month period: February to July 2011
• Ratio of recommendations clicked/menus showing recommendations is 2.5:1
• Recommendations are shown for an average of 30% of menus in OCUL
• Clickthrough Rate of recommendations followed varies a lot by school. Waterloo is lower at around 3% while some others are over 10%
OCUL Usage Data Findings (2)

• Relationships? Too early to tell, but for now...
  – ILL is not increasing
  – Get holding service seems to be decreasing the most as clickthroughs rise for recommendations
  – ILL requests from bx as the source is low
  – ILL requests from SFX were 22% of the total ILL borrowing and almost 50% of journal article requesting in 2010. This has risen steadily since it became available in 2007/8.
Focus Groups - Waterloo

• Questions derived from previous evaluations (Kansas, CDL, Guelph)
• Primo context for recommendations
bX Recommender @ University of Waterloo Library
bX Recommender @ University of Waterloo Library

• June 10, 2011 bX Recommender activated in Waterloo’s Primo Central trial view
bX Recommender @ University of Waterloo Library

• Assessment Goal – learn more about user preferences and success in using Primo Central - including bX Recommender
bX Recommender @ University of Waterloo Library

- bX Recommender early findings

- During the period from Aug. 22 – Sept 2., 12 students (8 undergrad, 4 grad) were interviewed
bX Recommender @ University of Waterloo Library

• Goal – to determine:
  1. What students think of recommender services
  2. What students think of actual bX Recommendations
  3. Get some feedback on the “usability” of the service
bX Recommender @ University of Waterloo Library

• Findings:

  – Overall, the students are generally positive about services that provide recommendations
bX Recommender @ University of Waterloo Library

• Findings:

  – “It’s good to know what other people found and actually used. It can save you a lot of time searching”
bX Recommender @ University of Waterloo Library

• Findings:

  1. high expectations that the recommendations will yield useful information
bX Recommender @ University of Waterloo Library

• Findings:

1. high expectations that the recommendations will yield useful information
2. there is an expectation that recommended articles will be available online
bX Recommender @ University of Waterloo Library

• Findings:

1. high expectations that the recommendations will yield useful information
2. there is an expectation that recommended articles will be available online
3. if the recommendations are not useful, students will abandon use of the service
bX Recommender @ University of Waterloo Library

• Findings:

– When performing a search for “global warming” users generally thought article recommendations were relevant
bX Recommender @ University of Waterloo Library

• Findings:

  – When performing a search for “global warming” users generally thought article recommendations were relevant
  – However, they were quick to pick up on the ones that seemed out of place
bX Recommender @ University of Waterloo Library

• Findings:

  – Students were asked if they thought they might have found these articles if they had not been listed as a recommendation
bX Recommender @ University of Waterloo Library

• Findings:

  – Overall usability (design, ease of use, how information is presented & organized)
bX Recommender @ University of Waterloo Library

• Findings:

  – Students thought it was “pretty straightforward”, “easy to use”, “convenient”
bX Recommender @ University of Waterloo Library

• Findings:

  – Students thought recommendations were “hidden”
Findings:

- There was some confusion about the different citation styles that students encountered when they looked at the recommendations.
bX Recommender @ University of Waterloo Library

- Findings:

<table>
<thead>
<tr>
<th>#</th>
<th>Authors</th>
<th>Title</th>
<th>Journal/Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BROECKER, Wallace S.</td>
<td>CLIMATIC CHANGE - ARE WE ON BRINK OF A PRONOUNCED GLOBAL WARMING</td>
<td>Science 189.4201 (1975) 460-463</td>
</tr>
<tr>
<td>3</td>
<td>Kuzyakov, Yakov</td>
<td>Priming effects: Interactions between living and dead organic matter</td>
<td>Soil biology &amp; biochemistry 42.9 (2010) 1363-1371</td>
</tr>
<tr>
<td>4</td>
<td>ONISHI, Akira</td>
<td>The impact of CO2 emissions on the world economy Policy simulations of FUGI global model</td>
<td>Journal of policy modeling 29.6 (2007) 797-819</td>
</tr>
</tbody>
</table>
bX Recommender @ University of Waterloo Library

• Findings:

  – Thumbs up/thumbs down – how does this work?
    • “Does that get saved to your user account, like if you like it or don’t like it?”
    • It would be nice to see “numbers like 3/5 people liked this article”
bX Recommender @ University of Waterloo Library

• “I think it’s fantastic. I’m a big fan of recommendations ... if nothing else, it makes research easier ... I might not think of all the things that other people thought of that did similar research and found similar things. So it’s nice to kind of have the Internet hive mind working to my advantage”
Next Steps – OCUL and Institutional Levels

• Work with the bx api
• Analyze longer-term usage stats
• Contribute data and assess whether this improves our recommendations
• Get feedback from undergraduate students
References


