Personalizing Web Search using Long Term Browsing History

Nicolaas Matthijs (University of Cambridge, UK)
Filip Radlinski (Microsoft, Vancouver)

WSDM 2011
What is personalized web search?
What is personalized web search?

Present each user a different ranking tailored to their personal interests and information need

= Personalized web search
Many approaches

Clickthrough-based approaches
- PClick (Dou et al., 2007)
- Promote URLs previously clicked by the same user for the same query

Profile-based approaches
- Teevan et al., 2005
- Rich model of user interests
- Built from search-related information, previously visited web sites, documents on hard-drive, e-mails, etc.
- Re-rank top returned search results
Goal

- Improve on existing personalized web search techniques
  - Combine a profile-based approach with a clickthrough-based approach
  - Selection of new features
  - Build an improved user representation from long term browsing history
- Improve on the evaluation methodology
  - Find out whether search personalization makes a difference in real life
  - Improve search result ranking without changing search environment
- Develop tool used by real people
Search Personalization Process

1) User Interest Extraction
   • User Profile representing user’s interests
     ✓ List of weighted terms
   • List of all visited URLs and number of visits
   • List of all search queries and results clicked

2) Result re-ranking
   • Change order of results to better reflect user’s interests
   • Get first 50 results for query on Google
   • Re-rank based on user profile by giving a score to each snippet
User Profile Extraction

Step 1: Term List Generation

- Don’t treat web pages as normal flat documents but as structured documents
- Use different sources of input data
  - Title unigrams
  - Metadata description unigrams
  - Metadata keywords
  - Full text unigrams
  - Extracted terms (Vu et al., 2008)
  - Extracted noun phrases (Clark et al., 2007)
- Specify how important each data source is (weight vector)
- Combination of data sources

=> List of terms to be associated with the user
User Profile Extraction

Step 2: Term List Filtering

- No filtering
- WordNet based POS filtering
- Google N-Gram corpus based filtering

=> Filtered list of terms
Step 3: Term Weighting

- **TF**
  \[ \vec{F}_{ti} = \begin{bmatrix} f_{title_{ti}} \\ f_{mdesc_{ti}} \\ f_{text_{ti}} \\ f_{mkeyw_{ti}} \\ f_{terms_{ti}} \\ f_{nphrases_{ti}} \end{bmatrix} \]
  \[ w_{TF}(t_i) = \vec{F}_{ti} \cdot \vec{\alpha} \]

- **TF-IDF**
  \[ w_{TFIDF}(t_i) = \frac{1}{\log(DF_{ti})} \times w_{TF}(t_i) \]

- **pBM25 (Teevan et al., 2005)**
  \[ w_{pBM25}(t_i) = \log \left( \frac{r_{ti} + 0.5}{n_{ti} + 0.5} \right) \left( \frac{N - n_{ti} + 0.5}{R - r_{ti} + 0.5} \right) \]

= User Profile: list of terms and term weights
Search Personalization Process

1) User Interest Extraction
   • User Profile representing user’s interests
     • List of weighted terms
     • List of all visited URLs and number of visits
     • List of all search queries and results clicked

2) Result re-ranking
   • Change order of results to better reflect user’s interests
   • Get first 50 results for query on Google
   • Re-rank based on user profile by giving a score to each snippet
Result re-ranking

Step 1: Snippet scoring

- Matching
  \[ \text{score}_M(s_t) = \sum_{z=1}^{N_s} f_z \times w(t_z) \]

- Unique Matching
  \[ \text{score}_{UM}(s_t) = \sum_{z=1}^{N_s} w(t_z) \]

- Language Model
  \[ \text{score}_{LM}(s_t) = \sum_{z=0}^{N_s} \log \left( \frac{w(t_z) + 1}{w_{total}} \right) \]

Step 2: Keep Google rank into account

Step 3: Give extra weight to previously visited pages
Evaluation

• Difficult problem
• Most previous work
  ▪ Small number of users evaluating relevance of small number of queries (Teevan et al., 2005)
  ▪ Simulate personalized search setting using TREC query and document collection
  ▪ After-the-fact log based analysis (Dou et al., 2007)
• Wanted to find out whether it yields a real difference in real-life usage
• Ideally: real-life usage data from lots of users over long time
• Unfeasible: high number of parameters

=> 2 step evaluation process
Evaluation: Capturing Data

• Need users and data to work with
• Full browsing history
• Not publicly available

⇒ Firefox add-on

AlterEgo Search 0.2
AlterEgo Search personalization based on your browsing history

• 41 users / 3 months
• 530,334 page visits / 39,838 Google searches
Step 1: Offline Relevance Judgments

- Identify most promising parameter configurations
- Offline evaluation session
- 6 users assess the relevance of the top 50 results for 12 queries
- Assess all possible combinations of all parameters
- Calculate NDCG score for each ranking
  (Jarvelin et al., 2000)

\[
NDCG@10 = \frac{1}{Z} \sum_{i=1}^{10} \frac{2^{rel_i} - 1}{\log_2(1 + i)}
\]
Step 1: Results

- 15,878 profile + re-ranking combinations investigated
- Compared to 3 baseline systems (Google, PClick and Teevan)
- 4,455 better than Google | 3,335 better than Teevan | 1,580 better than Pclick
- Identified 4 most promising personalization approaches
Step 1: Results

• Treating web pages as a flat document does not work.
• Advanced NLP techniques and keyword focused approaches work best.
• One re-ranking method outperforms all of the other ones:
  ✓ LM
  ✓ extra weight to visited URLs
  ✓ keeping the Google rank into account
Step 2: Online Interleaved Evaluation

- Assess the selected personalization techniques
- Extend Firefox add-on to do personalization in user’s browser as they go
- Interleaved evaluation using Team-Draft Interleaving algorithm (Radlinski et al., 2008)
- Shown to accurately reflect differences in ranking relevance (Radlinski et al., 2010)
Step 2: Online Interleaved Evaluation

Original ranking (Google)
1. Infrared - Wikipedia
   http://wikipedia.org/infrared
2. IRTech - Infrared technologies
   http://www.irtech.org
3. International Rectifier - Stock Quotes
   http://finance.yahoo.co.uk/IRE
4. SIGIR - New York Conference
   http://www.sigir.org
5. About Us - International Rectifier
   http://www.inrect.com

Personalized ranking
1. SIGIR - New York Conference
   http://www.sigir.org
2. Information Retrieval - Wikipedia
   http://wikipedia.org/ir
3. IRTech - Infrared technologies
   http://www.irtech.org
   http://wikipedia.org/infrared
5. About Us - International Rectifier
   http://www.inrect.com

Interleaved Ranking
1. SIGIR - New York Conference
   http://www.sigir.org (P)
2. Infrared - Wikipedia
   http://wikipedia.org/infrared (O)
3. IRTech - Infrared technologies
   http://www.irtech.org (O)
   http://wikipedia.org/ir (P)
5. International Rectifier
   http://www.inrect.com

Count which ranking is clicked most often
http://finance.yahoo.co.uk/IRE (O)
Results

• 41 users / 2 weeks / 7,997 queries

• MaxNDCG significantly (p < 0.001) outperforms Google
  MaxBestPar significantly (p < 0.01) outperforms Google
  MaxQuer significantly (p < 0.05) outperforms Google

• Run on all queries: 70% of queries untouched, 20% improved, 10% worse
  Average improvement of 4 ranks. Average deterioration of 1 rank.

• One strategy is consistently the best:
  TF-IDF, RTitle, RMKeyw, RCCParse, NoFilt - LM, Look At Rank, Visited
Future Work

• Expand set of parameters
  ✓ Learning optimal weight vector
  ✓ Using other fields

• Temporal information
  ✓ How much browsing history should be used?
  ✓ Decay weighting of older items
  ✓ Page visit duration

• Other behavioral information

• Use extracted profile for other purposes
Conclusion

• Outperform Google and previous best personalization strategies

• Build an improved user profile for personalization
  ✓ Not treat web pages as flat documents
  ✓ Use more advanced NLP techniques

• Improve upon the evaluation methodology
  ✓ First large online comparative evaluation of personalization techniques
  ✓ Investigate whether personalization makes difference in real life usage
  ✓ Done in academic setting, no large datasets available

• Tool that can be downloaded and used by everyone
  ✓ Code is open sourced, very clean and readable