Webis at the TREC 2012 Session track

Matthias Hagen   Martin Potthast   Matthias Busse   Jakob Gomoll
Jannis Harder   Benno Stein

Bauhaus-Universität Weimar
matthias.hagen@uni-weimar.de

TREC 2012
Gaithersburg
November 9, 2012
Two research questions . . .
Question 1: query expansion depending on session type

“Low risk” session

- QE might be beneficial
- Low risk of misunderstanding
Question 1: query expansion depending on session type

“Low risk” session
- QE might be beneficial
- Low risk of misunderstanding

“High risk” session
- QE considered harmful
- High risk of misunderstanding
Question 2: knowledge from other users’ sessions

Sessions with same goals
Two standard retrieval models

[chatnoir.webis.de]  [boston.lti.cs.cmu.edu/Services/]

- BM25F + PageRank + Proximity
- Used in runs 1 and 3

- Language modeling + inference network
- Used in run 2
Runs 1 and 2: query expansion by session types

Compare current query $q$ to each previous query

If $q$ is not a repetition, generalization, or specialization, then populate

$Q$: previous queries

$R$: previous results (documents)

$S$: previous snippets

$T$: previous titles

Query expansion approach

RL2: at most two keyphrases from $Q$

RL3: additionally at most one keyphrase from each $R$, $S$, $T$

RL4: only clicked results in $R$, $S$, $T$

Weights: 2.0 from $q$, 0.6 from $Q$, 0.2 from $R$, 0.1 from $S$ or $T$
Runs 1 and 2: query expansion by session types

Compare current query $q$ to each previous query

If $q$ is not a repetition, generalization, or specialization, then populate

- $Q$: previous queries
- $R$: previous results (documents)
- $S$: previous snippets
- $T$: previous titles

Query expansion approach

- RL2: at most two keyphrases from $Q$
- RL3: additionally at most one keyphrase from each $R$, $S$, $T$
- RL4: only clicked results in $R$, $S$, $T$

Weights: 2.0 from $q$, 0.6 from $Q$, 0.2 from $R$, 0.1 from $S$ or $T$
Runs 1 and 2: postprocessing

Result list postprocessing

- Aspect sessions: show Wikipedia
- VIP segments: find long Wikipedia title in $q$, show article
- Clicks: results from similar sessions at rank 3 and 4
- Long documents: remove when $\geq 7000$ words
- Duplicates: remove when 5-gram cosine similarity $\geq 0.98$

Run 2

- Indri instead of ChatNoir
- Query segmentation [Hagen et al., CIKM 2012]
### Runs 1 and 2: postprocessing

#### Result list postprocessing

- **Aspect sessions:** show Wikipedia
- **VIP segments:** find long Wikipedia title in \( q \), show article
- **Clicks:** results from similar sessions at rank 3 and 4
- **Long documents:** remove when \( \geq 7000 \) words
- **Duplicates:** remove when 5-gram cosine similarity \( \geq 0.98 \)

#### Run 2

- **Indri** instead of **ChatNoir**
- **Query segmentation** [Hagen et al., CIKM 2012]
### Runs 1 and 2: nDCG@10 influence

<table>
<thead>
<tr>
<th></th>
<th>RL1</th>
<th>RL2</th>
<th>RL3</th>
<th>RL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>run 1  (ChatNoir)</td>
<td>0.0865</td>
<td>0.1174↑</td>
<td>0.1204↑</td>
<td>0.1171↑</td>
</tr>
<tr>
<td>run 2  (Indri)</td>
<td>0.2053</td>
<td>0.2097↑</td>
<td>0.2102↑</td>
<td>0.2077↑</td>
</tr>
</tbody>
</table>

#### Observations
- ChatNoir’s initial performance rather low
- ChatNoir (BM25F) significantly benefits from risk-aware QE
- Indri (LM) benefits (not statistically significant)
Run 3: knowledge from other users’ sessions

Search shortcuts

- Query expansion with terms from related sessions
- RGU-ISTI-Essex team used Microsoft RFP 2006 log
- Performance gain not significant
- Not many related sessions found?!

[Baraglia et al., RecSys 2009]

Our idea

- Use TREC sessions as source, and
- Manual creation of more related sessions
  (three for sessions 1, 3, 8, 34, 38, 46, 53, 64, 66, 69, and 92)
- Should count as manual run?!
Run 3: knowledge from other users’ sessions

<table>
<thead>
<tr>
<th>Search shortcuts</th>
<th>[Baraglia et al., RecSys 2009]</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Query expansion with terms from related sessions</td>
<td></td>
</tr>
<tr>
<td>- RGU-ISTI-Essex team used Microsoft RFP 2006 log</td>
<td></td>
</tr>
<tr>
<td>- Performance gain not significant</td>
<td></td>
</tr>
<tr>
<td>- Not many related sessions found?!</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Our idea</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Use TREC sessions as source, and</td>
</tr>
<tr>
<td>- Manual creation of more related sessions (three for sessions 1, 3, 8, 34, 38, 46, 53, 64, 66, 69, and 92)</td>
</tr>
<tr>
<td>- Should count as manual run?!</td>
</tr>
</tbody>
</table>
Run 3: query expansion + postprocessing

Query expansion
- Analogous to runs 1 and 2, but
- \( Q, R, S, \) and \( T \) populated from related sessions only

Result list postprocessing
- Analogous to runs 1 and 2, but
- Top ranks populated with clicks from related sessions only
Run 3: nDCG@10 influence

<table>
<thead>
<tr>
<th></th>
<th>RL1</th>
<th>RL2</th>
<th>RL3</th>
<th>RL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>run 1 (same session)</td>
<td>0.0865</td>
<td>0.1174</td>
<td>0.1204</td>
<td>0.1171</td>
</tr>
<tr>
<td>run 3 (other sessions)</td>
<td>0.1086</td>
<td>0.1220</td>
<td>0.1401</td>
<td>0.1796</td>
</tr>
</tbody>
</table>

Observations

- Other users’ sessions can help a lot (risk-aware)
- More than the same users’ previous interactions
Run 3: the best from both worlds?! 

Low risk + related sessions
Almost the end: The take-home messages!
What we have done

Main results

- Risk-aware session type consideration
  - Mostly performance gains, hardly any losses

- Impact on standard retrieval models
  - $BM25F \uparrow$ vs. $Indri \uparrow$

- Other users’ sessions
  - 65% improvement for $BM25F$

Future work

- More fine-grained types
- Other retrieval models
- QE techniques
- When to step in?
What we have (not) done

Main results

- Risk-aware session type consideration
  \(\text{\(\hookrightarrow\)}\) mostly performance gains, hardly any losses

- Impact on standard retrieval models
  \(\text{\(\hookrightarrow\)}\) \(BM25F\) \(\uparrow\) vs. \(Indri\) \(\uparrow\)

- Other users’ sessions
  \(\text{\(\hookrightarrow\)}\) 65% improvement for \(BM25F\)

Future work

- More fine-grained types
- Other retrieval models
- QE techniques
- When to step in?
What we have (not) done

Main results

- Risk-aware session type consideration
  \(\rightarrow\) *mostly performance gains, hardly any losses*

- Impact on standard retrieval models
  \(\rightarrow\) *BM25F \(\uparrow\) vs. Indri \(\uparrow\)*

- Other users’ sessions
  \(\rightarrow\) *65% improvement for BM25F*

Future work

- More fine-grained types
- Other retrieval models
- QE techniques
- When to step in?

Thank you

Hagen et al.  Webis at the TREC 2012 Session track