Learning Ranking Functions with SVMs

CS4780/5780 – Machine Learning
Fall 2014
Thorsten Joachims
Cornell University

http://www.cs.cornell.edu/People/tj/publications/joachims_02c.pdf

Adaptive Search Engines

- Traditional Search Engines
  - One-size-fits-all
  - Hand-tuned retrieval function
- Hypothesis
  - Different users need different retrieval functions
  - Different collections need different retrieval functions
- Machine Learning
  - Learn improved retrieval functions
  - User Feedback as training data

Overview

- How can we get training data for learning improved retrieval functions?
  - Explicit vs. implicit feedback
  - Absolute vs. relative feedback
  - User study with eye-tracking and relevance judgments
- What learning algorithms can use this training data?
  - Ranking Support Vector Machine
  - User study with meta-search engine

Sources of Feedback

- Explicit Feedback
  - Overhead for user
  - Only few users give feedback
  => not representative
- Implicit Feedback
  - Queries, clicks, time, mousing, scrolling, etc.
  - No Overhead
  - More difficult to interpret

Feedback from Clickthrough Data

| Relative Feedback: Clicks reflect preference between observed links. |
| Absolute Feedback: The clicked links are relevant to the query. |

- (3 < 2), Rel(1), NotRel(2), Rel(3), NotRel(4), NotRel(5), NotRel(6), Rel(7)
- (7 < 2), Rel(1), NotRel(2), Rel(3), NotRel(4), NotRel(5), NotRel(6), Rel(7)
- (7 < 4), Rel(1), NotRel(2), Rel(3), NotRel(4), NotRel(5), NotRel(6), Rel(7)
- (7 < 6), Rel(1), NotRel(2), Rel(3), NotRel(4), NotRel(5), NotRel(6), Rel(7)

User Study: Eye-Tracking and Relevance

- Scenario
  - WWW search
  - Google search engine
  - Subjects were not restricted
  - Answer 10 questions
- Eye-Tracking
  - Record the sequence of eye movements
  - Analyze how users scan the results page of Google
- Relevance Judgments
  - Ask relevance judges to explicitly judge the relevance of all pages encountered
  - Compare implicit feedback from clicks to explicit judgments

6. Pages of SUPPORT VECTOR MACHINES... http://www.jiscmail.ac.uk/lists/SUPPORT...
What is Eye-Tracking?
Device to detect and record where and what people look at
- **Fixations**: ~200-300ms; information is acquired
- **Saccades**: extremely rapid movements between fixations
- **Pupil dilation**: size of pupil indicates interest, arousal

How Many Links do Users View?
![Bar chart showing the total number of abstracts viewed per page with mean 3.07, median/mode 2.00](chart.png)

In Which Order are the Results Viewed?
![Graph showing the instance of arrival to each result with users tending to read in order](graph.png)

Looking vs. Clicking
![Graph showing the mean time spent in each abstract with users viewing links one and two more thoroughly/often and clicking most frequently on link one](chart2.png)

Do Users Look Below the Clicked Link?
<table>
<thead>
<tr>
<th>Viewed Rank</th>
<th>Viewed &amp; Clicked</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>90.6%</td>
</tr>
<tr>
<td>2</td>
<td>56.8%</td>
</tr>
<tr>
<td>3</td>
<td>30.2%</td>
</tr>
<tr>
<td>4</td>
<td>17.3%</td>
</tr>
<tr>
<td>5</td>
<td>8.6%</td>
</tr>
<tr>
<td>6</td>
<td>4.3%</td>
</tr>
</tbody>
</table>

=> Users typically do not look at links below before they click (except maybe the next link)

How do Clicks Relate to Relevance?
- **Experiment (Phase II)**
  - Additional 16 subjects
  - Manually judged relevance
    - Abstract
    - Page
- **Manipulated Rankings**
  - **Normal**: Google’s ordering
  - **Swapped**: Top Two Swapped
  - **Reversed**: Ranking reversed
- **Experiment Setup**
  - Same as Phase I
  - Manipulations not detectable

1. [Kernel Machines](http://www.kernel-machines.org/)
2. [Support Vector Machine](http://jbolivar.freeservers.com/)
3. [SVM-Light Support Vector Machine](http://ais.gmd.de/~thorsten/svm_light/)
4. [An Introduction to SVMs](http://www.support-vector.net/)
6. [Archives of SUPPORT VECTOR ...](http://www.jisc.ac.uk/lists/SUPPORT...)
8. [Royal Holloway SVM](http://svm.dcs.rhbnc.ac.uk)
9. [SVM World](http://www.svmworld.com)
10. [Fraunhofer FIRST SVM page](http://svm.first.gmd.de)
Presentation Bias

Hypothesis: Order of presentation influences where users look, but not where they click!

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Rank of clicked link as sorted by relevance judges</th>
</tr>
</thead>
<tbody>
<tr>
<td>“normal”</td>
<td>Normal + Swapped: 2.67</td>
</tr>
<tr>
<td>“swapped”</td>
<td>=&gt; Users click on less relevant links, if they are embedded between irrelevant links.</td>
</tr>
</tbody>
</table>

Are Clicks Absolute Relevance Judgments?

- Clicks depend not only on relevance of a link, but also
  - On the position in which the link was presented
  - The quality of the other links
- Interpreting Clicks as absolute feedback extremely difficult!

Comparison with Explicit Feedback

<table>
<thead>
<tr>
<th>Explicit Feedback Data Strategy</th>
<th>Abstracts Phase I “normal”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-Judge Agreement Click &gt; Skip Above</td>
<td>89.5</td>
</tr>
<tr>
<td>Last Click &gt; Skip Above</td>
<td>80.8 ± 3.6</td>
</tr>
<tr>
<td>Click &gt; Earlier Click</td>
<td>83.1 ± 3.8</td>
</tr>
<tr>
<td>Click &gt; Skip Previous</td>
<td>67.2 ± 12.3</td>
</tr>
<tr>
<td>Click &gt; No Click Next</td>
<td>82.3 ± 7.3</td>
</tr>
<tr>
<td>=&gt; All but “Click &gt; Earlier Click” appear accurate</td>
<td></td>
</tr>
</tbody>
</table>

Strategies for Generating Relative Feedback

- “Click > Skip Above”
  - (3>2), (5>2), (5>4)
- “Last Click > Skip Above”
  - (5>2), (5>4)
- “Click > Earlier Click”
  - (3>1), (5>1), (5>3)
- “Click > Skip Previous”
  - (3>2), (5>4)
- “Click > Skip Next”
  - (1>2), (3>4), (5>6)

Is Relative Feedback Affected by Bias?

<table>
<thead>
<tr>
<th>Explicit Feedback Data Strategy</th>
<th>Abstracts Phase II “normal”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Click &gt; Skip Above</td>
<td>88.0 ± 9.5</td>
</tr>
<tr>
<td>Last Click &gt; Skip Above</td>
<td>89.7 ± 9.8</td>
</tr>
<tr>
<td>Click &gt; Earlier Click</td>
<td>75.0 ± 25.8</td>
</tr>
<tr>
<td>Click &gt; Skip Previous</td>
<td>88.9 ± 24.1</td>
</tr>
<tr>
<td>Click &gt; No Click Next</td>
<td>75.6 ± 14.5</td>
</tr>
<tr>
<td>=&gt; Significantly better than random in all conditions, except “Click &gt; Earlier Click”</td>
<td></td>
</tr>
</tbody>
</table>
Learning Retrieval Functions from Pairwise Preferences

- Idea: Learn a ranking function, so that number of violated pair-wise training preferences is minimized.
- Form of Ranking Function: sort by
  \[ U(q,d) = w_1 \cdot \text{(#of query words in title of } d) + w_2 \cdot \text{(#of query words in anchor)} + \ldots + w_r \cdot \text{(page-rank of} d) = w \cdot \Phi(q,d) \]
- Training: Select \( w \) so that
  \[ \text{if user prefers} \ d_i \text{ to} \ d_j \text{ for query} \ q, \ \text{then} \ U(q,d_i) > U(q,d_j) \]

Experiment

- Meta-Search Engine “Striver”
  - Implemented meta-search engine on top of Google, MSNSearch, Altavista, Hotbot, Excite
  - Retrieve top 100 results from each search engine
  - Re-rank results with learned ranking functions
- Experiment Setup
  - User study on group of ~20 German machine learning researchers and students
  - homogeneous group of users
  - Asked users to use the system like any other search engine
  - Train ranking SVM on 3 weeks of clickthrough data
  - Test on 2 following weeks

Results

<table>
<thead>
<tr>
<th>Ranking A</th>
<th>Ranking B</th>
<th>A better</th>
<th>B better</th>
<th>Tie</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learned</td>
<td>Google</td>
<td>29</td>
<td>13</td>
<td>27</td>
<td>69</td>
</tr>
<tr>
<td>Learned</td>
<td>MSNSearch</td>
<td>18</td>
<td>4</td>
<td>7</td>
<td>29</td>
</tr>
<tr>
<td>Learned</td>
<td>Toprank</td>
<td>21</td>
<td>9</td>
<td>11</td>
<td>41</td>
</tr>
</tbody>
</table>

Result:
- Learned > Google
- Learned > MSNSearch
- Learned > Toprank

Toprank: rank by increasing minimum rank over all 5 search engines
Learned Weights

- **Weight** | Feature
- 0.60 | cosine between query and abstract
- 0.48 | ranked in top 10 from Google
- 0.24 | cosine between query and the words in the URL
- 0.24 | doc ranked at rank 1 by exactly one of the 5 engines
- ... | ... (continue with more features)
- 0.22 | host has the name "citeseer"
- ... | ... (continue with more features)
- 0.17 | country code of URL is ".de"
- 0.16 | ranked top 1 by HotBot
- ... | ... (continue with more features)
- -0.15 | country code of URL is ".fi"
- -0.17 | length of URL in characters
- -0.32 | not ranked in top 10 by any of the 5 search engines
- -0.38 | not ranked top 1 by any of the 5 search engines

Conclusions

- Clickthrough data can provide accurate feedback
  - Clickthrough provides relative instead of absolute judgments
- Ranking SVM can learn effectively from relative preferences
  - Improved retrieval through personalization in meta search
- Current and future work
  - Exploiting query chains
  - Other implicit feedback signals
  - Adapting intranet search for ArXiv.org
  - Recommendation
  - Robustness to "click-spam"
  - Learning and micro-economic theory for interactive learning with preference
  - Further user studies to get better models of user behavior

Feedback across Query Chains