Supervised Reranking for Web Image Search

Linjun Yang\textsuperscript{1} and Alan Hanjalic\textsuperscript{2}

\textsuperscript{1} Microsoft Research Asia, China
\textsuperscript{2} Delft University of Technology, The Netherlands

Current Web Image Search Ranking

- The images are indexed mainly using the text in the associated web pages
- Feature-based ranking model

Problem Cases (1)

Example Query: Red Wine

(a) Missing

(b) Mismatch

Problem Cases (2)

Example Query: George W. Bush

(a) Mismatch between the image content and the words

(b) These images, though all relevant, are hard to differentiate and rank based on text

Content-aware Image Search

- Problem Summary
  - The associated text is not the essential description of images, and tends to be mismatching.

- Existing Solutions
  - Concept-based image search
  - Visual reranking

Concept-based Image Search

(a.k.a. Semantic Indexing)

Linking the query with semantic concept detectors (e.g. LSCOM)

- Bridging the semantic gap
  - Scalability: It requires a large amount of labeled images to train concept models
  - Index capability: Indexing images as concepts leads to information loss
Visual Reranking

- Reorder the text-based ranking list with visual consistency
  - Based on initial text-based search results
  - Exploit the intrinsic consistency

+ Lightweight, unsupervised, and scaled-up
+ Hard to trade-off the contribution from the textual and visual domains
- The reranking model is hand-designed, which may not capture users’ true needs

Supervised Reranking

- Challenging problem
  - How to extract ranking features for a pair of textual query and visual document?

- Solution
  - Reranking features
    - Textual domain
      - Based on the initial text-based ranking
    - Visual domain
      - The textual query as the ranked images in the initial result
  - Learning
    - Adapted Ranking SVM

Comparison of Existing Approaches

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalability</td>
<td>Challenging</td>
</tr>
<tr>
<td>Index capability</td>
<td>Weak (concept)</td>
</tr>
<tr>
<td>Model representability</td>
<td>Good</td>
</tr>
</tbody>
</table>

Comparison of Methods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalability</td>
<td>Challenging</td>
<td>Easy to scale up</td>
</tr>
<tr>
<td>Index capability</td>
<td>Weak (concept)</td>
<td>Good (visual feature)</td>
</tr>
<tr>
<td>Model representability</td>
<td>Good</td>
<td>Weak</td>
</tr>
</tbody>
</table>

Feature Design Principles

- Visual consistency
  - Visually similar images should have similar ranks

- Pseudo Relevance feedback
  - Top ranked images in the text-based search are relevant

Data set

- collected top 1000 images with 29 queries from 3 image search engines
Reranking Features

- Visual consistency
  - Nearest neighborhood
  - Reverse neighborhood
- Pseudo-relevance feedback
  - Models from top results
- Initial ranking based feature

<table>
<thead>
<tr>
<th>Features</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVI</td>
<td>Hard Voting of Neighbors</td>
</tr>
<tr>
<td>RNI</td>
<td>Initial Rank-based Soft Voting of Neighbors</td>
</tr>
<tr>
<td>RI</td>
<td>Initial Rank Voting of Reciprocal Neighbors</td>
</tr>
<tr>
<td>RS</td>
<td>Initial Rank-based Soft Voting of Reciprocal Neighbors</td>
</tr>
<tr>
<td>SAVI</td>
<td>Neighbor Rank-based Soft Voting of Reciprocal Neighbors</td>
</tr>
<tr>
<td>SAVN</td>
<td>Neighbor Rank-based Soft Voting of Neighbors</td>
</tr>
<tr>
<td>PLEX</td>
<td>Local Density Estimation for PLEX</td>
</tr>
<tr>
<td>SREP</td>
<td>Soft Relevance Estimation for PLEX</td>
</tr>
<tr>
<td>IR</td>
<td>Initial Ranking</td>
</tr>
</tbody>
</table>

Think about it Broadly...

- The origin of the scalability
  - Embedding the queries into features
    - Feature extracted from <query, document> pairs
    - The model to be learned is independent of the queries
- Related schemes
  - Learning to rank
    - Features extracted from <query, text document> pair
    - tf-idf, BM25, LMR, etc.
  - Distance metric learning
    - Features extracted from <image, image> pair
  - Query-relative Learning

Performance Comparison (1)

- Supervised reranking outperforms the other unsupervised ones and text baseline
- The superiority of supervised reranking over Bayesian reranking is statistically significant

Performance Comparison (2)

- MSRA-MM dataset1
  - 68 queries, 1000 top images, from Bing
  - Evaluation measure NDCG

Individual Performance of Features

- The features have reasonable performance on their own.

Leave One Feature Out

- Leaving some feature out improves the performance
  - Due to the correlation among features, good features may not contribute positively

Leave One Feature Out

- Leaving some feature out improves the performance
  - Due to the correlation among features, good features may not contribute positively
  - It suggests automatic feature selection

Adapted Ranking SVM

- Adapted Ranking SVM performs better than standard Ranking SVM (alpha=1).
- Setting alpha = # of visual features achieves the best result.

Reranking Results (query: George W. Bush)

Text-based

Reranking

Conclusion and Future Work

- Conclusion
  - We proposed a new scheme, supervised reranking, for web image search.
  - Supervised reranking achieves both effectiveness and scalability by learning a query-independent model.
- Future work
  - Automatic feature selection for supervised reranking.
  - Supervised ranking with textual and visual ranking features.
Thanks !