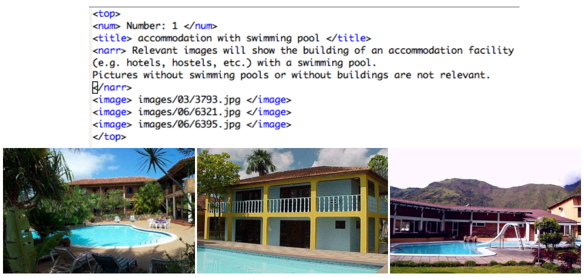
The Importance of the Depth for Text-Image Selection Strategy in Learning-To-Rank

David Buffoni, Sabrina Tollari, Patrick Gallinari
Université Pierre et Marie Curie -Paris 6, UMR CNRS 7606 - LIP6, Paris, France, firstname.lastname@lip6.fr

Context: Text-Image Retrieval



Task: produce a sorted list of images given a user query

Problem: how to efficiently learn a ranking function ?

Our work: evaluation of the impact of the depth of a pooling methodology on Learning-to-Rank (LTR) algorithms.

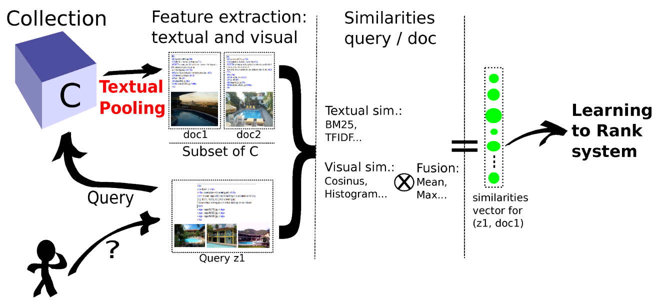
Pooling Methodology

Aim: improve diversity and avoid being biased by only one similarity during the learning step.

Depth-k pooling:

- build lists of top-k documents retrieved by models (BM25, TFIDF and 2 language models)
- merge the lists to obtain a training set

Our work: what is the influence of k on LTR algorithms in text-image retrieval ?



Baseline and Learning-To-Rank Models

Baseline: $BL(q,d) = \lambda S_{Text}(q_t,d_t) + (1-\lambda) S_{Visual}(q_v,d_v)$ where $S_{Text}(q_t,d_t) = BM25$ and $S_{Visual}(q_v,d_v) = max_{q_i}(Histo_{HSV}(q_i,d_i))$ is the maximum fusion operator applied of histogram distances of HSV.

SVM^{RANK} [2]: LTR algorithm formulated as a SVM problem and optimizing the mean rank of the relevant documents (the mean number of pairwise errors).

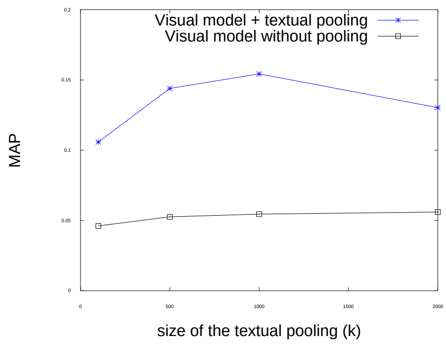
OWPC [3]: LTR algorithm which optimizes a loss function focused on the top of the list by fixing Ordered Weighted Averaging (OWA)[4] weights.

Experiments - Results

Collection and features:

- ImageCLEFphoto'06 (20,000 images with captions, 60 query topics with binary assessments)
- 60 extracted query/document similarities (BM25, SIFT....) normalized by query and by similarity
- 5 cross-validation folds where each fold contains training/validation/test query sets

Textual pooling influence on the visual model:



Textual pooling has a strong impact on the visual model ($BL_{\lambda=0}$). Consequently, constituting a filtered pool by textual information improves visual model performances.

Depth-k pooling on LTR algorithms:

k=	100	500	1000	2000
$BL_{\lambda=1}$ (BM25)	**0.240**	0.294	0.299	0.302
$BL_{\lambda=0}$ (visual)	0.106	0.144	0.154	0.130
$BL_{\lambda=*}$	0.235	0.278	0.285	0.291
SVM^{RANK}	0.220	0.283	0.297	0.294
$OWPC$	0.226	**0.297**	**0.306**	**0.303**

Conclusion

- OWPC gives the best results and SVM^{RANK} has surprisingly lower performances than BM25.
- Textual pooling improves models performances.
- The depth of the pooling (k) need to be well fixed.
- Future direction: effects of adding visual models in the pooling step.

References:

[1] J.A. Aslam et al. Document selection methodologies for efficient and effective learning-to-rank. In *SIGIR'09*.

[2] N.Usunier et al. Ranking with ordered weighted pairwise classification. In *ICML'09*.

[3] Thorsten Joachims. Optimizing search engines using clickthrough data. In *KDD'02*.

[4] R.R. Yager. On ordered weighted averaging aggregation operators in multi-criteria decision making. In *IEEE Trans. on Syst., Man and Cybernetics'88*.