Workshop of the Cross-Language Evaluation Forum for European Languages

CLEF 2009: Multilingual Information Access Evaluation II. Multimedia Experiments pp 150-156 | Cite as

Image Query Expansion Using Semantic Selectional Restrictions

Authors | Authors and affiliations

Osama El Demerdash, Sabine Bergler, Leila Kosseim

Part of the Lecture Notes in Computer Science book series (LNCS, volume 6242)

Abstract

This paper describes our participation at ImageCLEF 2009. We participated in the

Image Query Expansion Using Semantic Selectional Restrictions

Osama El Demerdash, Sabine Bergler and Leila Kosseim
Concordia University
{osama_el,bergler,kosseim}@cse.concordia.ca

CLaC Laboratory - Department of Computer Science & Software Engineering - Concordia University

Abstract. This paper describes our participation at ImageCLEF 2009. We participated in the photographic retrieval task (ImageCLEFPhoto). Our method is based on cross-media pseudo-relevance feedback. We have enhanced the pseudo-relevance feedback mechanism by using semantic selectional restrictions. We use Terrier for text retrieval and our own simple block-based visual retrieval engine. The results obtained at ImageCLEF 2009 show that our method is robust and promising. However, there is room for improvement on the visual retrieval.

1 Introduction

This paper describes our participation at ImageCLEF 2009. We participated in the photographic retrieval task (ImageCLEFPhoto). This year's task targeted the promotion of diversity in image search. It involved an annotated image collection of approximately half a million images, and fifty queries divided into two sets: one with a subject and provided specific subtopics (clusters), while the other with only a topic. A full description of the task can be found in [1].

We submitted four runs, aiming at evaluating our method as well as the resources used. Similar to our participation at ImageCLEF 2008[2], our method is based on cross-media pseudo-relevance feedback. However, in order to account for the much larger data set, we have introduced some modifications to our visual component. We have also enhanced the textual retrieval component, as well as the pseudo-relevance feedback mechanism by using semantic selectional restrictions.

The rest of this paper is organized as follows: Section 2 describes the visual retrieval component, Section 3 the text processing of the query, Section 4 the enhanced pseudo-relevance feedback and Section 5 the results we obtained at ImageCLEF 2009, then we conclude the paper.

2 Visual Retrieval

Figure 1 shows the different regional divisions used to analyze the image. In order to capture different levels of detail, we divide the image into 2X2, 3X3,

Fig. 1. Block-based Visual Features.

4X4 blocks yielding 4, 9, 16 equal partitions respectively. Due to the much larger size of the data set compared to the IAPR TC-12 collection (20,000 images) used in previous years, we resorted to reducing the index by eliminating some of the descriptors we used previously, such as the grey-level and gradient-magnitude descriptors. The image is first converted to the Intensity/Hue/Saturation (IHS) color space, a perceptual color space which is more intuitive and reflective of human color perception than the RGB color space. This also allows for assigning more weight to the hue component which is a better discriminating feature as shown in [3].

As previously illustrated [4], the moments of histograms are efficient approximations of the entire histogram. Therefore, for each of the three-band color histograms of the divisions, the first two moments (the mean and the average energy) as well as the standard deviation are stored in the index.

For retrieval, the different partitions are compared to their counter parts in the query images. We selected the Manhattan distance (L^1 Norm) after investigating several other measures including the Euclidean and the Mahalanobis distances, combined with a measure for the number of blocks within a minimum threshold for the distance. Since all features were represented as histograms with the same number of bins (256), no normalization was necessary. The images in the database were ranked according to their highest proximity to any of the three query images. This choice presumes that our simple features do not perform equally well on all example images.

3 Text Retrieval

The text is tokenized and preprocessed by removing stop words (grammatical words which do not contribute to the meaning) and punctuation. The rest of the terms are converted to lower case and stemmed using the Snowball stemmer [5].

Queries are tokenized and preprocessed similarly; stop words and punctuation are removed and the rest of the terms are stemmed. The queries consist of the topic and the cluster description (where available), in addition to the expansion terms from the top visual results. Named-entities, recognized using simple capitalization heuristics, are given more weight and multiple-token named-entities are chunked into one term by adding quotes around them.

For text retrieval, we use the Terrier Information Retrieval platform, a Java-based Information Retrieval platform available from the University of Glasgow [6]. Terrier includes boolean, vector-space and probabilistic model capabilities. We use the vector-space model, which slightly surpasses the probabilistic models in our experiments. In the vector-space model, documents and queries are represented as vectors of terms weighed by Term Frequencies multiplied by the Inverse Document Frequency (TF-IDF). Terrier also has the option of block-indexing for phrase querying which we employ. Query terms are considered unioned by Terrier in order to promote recall.

4 Pseudo-Relevance Feedback with Semantic Selectional Restrictions

In this phase, the query is first expanded with terms potentially related to the query, then sent to the text retrieval engine described in the previous section. Common ways for text query expansion include adding synonyms and other related terms to the query. However, according to our experiments, this approach leads to the introduction of many noisy terms. Instead, we opted for the extraction of related terms from the five highest-ranked results retrieved by the content-based system described in Section 2. For the data set in our experiments, all the terms associated with the image are extracted except for stop words. In order to expand the query without introducing noise, the candidate text is compared to the query topic. If the image is found to be potentially related to the topic, then the text query is expanded with the relevant terms and sent to the text retrieval engine. To compute the relatedness of the image annotation to the topic, we use the minimum threshold of one common non-grammatical word, due to data sparseness. An example of an exapnsion is the query "Obama", which was expanded with the following word stems: michell, wife, democrat, presidenti, candid, u s, senat, barack, democrat, illinoi, wave, introduc, ralli, univers, illinoi, chicago, pavilion, midst, offici, campaign, trip, iowa, new hampshir, formal, announc, candidaci, epa, tannen, mauri.

The purpose of the query expansion module is not only to augment the query by adding new candidate related terms to it, but also to enhance it by adding weights to its key terms and filtering out potentially noisy terms from expansion. We also avoid expanding the query with named entities that do not have a semantic relationship with the query. This is crucial in photographic collections, since by their nature, photographs and image queries are often bound by geographical constraints. In order to ensure that potential expansion images do not introduce conflicting geographical terms in the query, we first build a filter from the location specified in the query. We make use of WordNet [7], a lexical database, by traversing its *PartMeronym* hierarchy. A *PartMeronym* is a relationship between two nouns where the child noun constitutes a part of the parent noun. For geographical locations, this translates by the divisions of the parent noun. For example for the USA, a traversal of the hierarchy produces the names of the states, then major cities and towns followed by specific locations. While similar filters are possible for common nouns and using other relations such as Hyponymy (sub-classes of a term), we limit the expansion to named-entities, so as to avoid the problem of disambiguation of the specific sense of the term.

5 Results

Due to a glitch in the preprocessing of the differently formatted queries without given clusters, the official results we obtained at ImageCLEF 2009 do not reflect the actual performance of our methods. The official results can be found in [8] as well as the ImageCLEFphoto track description in this volume.

Table 1 shows the results we obtained when we fixed the glitch. The first two runs are purely visual and textual respectively. The *PRF* run combines visual and text retrieval using the Pseudo-relevance feedback mechanism described in section 4 and separate queries for each cluster, the results of which are then combined using a simple interleaving method. The *Combined* run uses the same method as the *PRF*, while combining all clusters information into one query. *P10* and *P20* are the Precision figures at 10 and 20 retrieved documents respectively. *CR10* and *CR20* are the Cluster Recall levels at 10 and 20 retrieved documents, while the F-measure reported in these tables employs P10 and CR10 similar to the official F-measure used at the 2009 ImageCLEF campaign.

Table 1. Results on ImageCLEFPhoto 2009 Queries.

Description	**P10**	**P20**	**CR10**	**CR20**	**MAP**	**Rel_Ret**	**F-measure**
Visual	0.0960	0.0990	0.2980	0.4340	0.0060	657	0.1452
Text	**0.7540**	**0.7800**	0.6877	0.7525	**0.4879**	**19148**	**0.7193**
With PRF	0.5820	0.6770	**0.7334**	**0.8482**	0.4221	17880	0.6490
Combined	0.6200	0.7090	0.6822	0.7972	0.4531	18387	0.6496

Our results show that using text only queries outperforms the pseudo-relevance feedback runs in F-measure as well as precision. However, the diversity of the pseudo-relevance feedback runs tends to be higher. The visual-only run rated very poorly. Indeed, the successful pseudo-relevance appeared to stem from expanding using the text associated with the example images, which were eliminated from the gold standard and did not count as valid results.

Tables 2 and 3 show the breakdown of these runs by query set (queries where the cluster information was given and queries without cluster information respectively).

Table 2. Queries with Given Clusters.

Description	**P10**	**P20**	**CR10**	**CR20**	**MAP**	**Rel_Ret**	**F-measure**
Visual	0.0720	0.0820	0.2603	0.3934	0.0026	241	0.1128
Text	**0.7400**	**0.7660**	**0.7796**	**0.8693**	**0.4595**	8778	**0.7593**
With PRF	0.5400	0.6900	0.7562	0.8772	0.4207	8664	0.6300
Combined	0.6000	0.7220	0.6741	0.7702	0.4476	**8793**	0.6349

Table 3. Queries without Given Clusters.

Description	**P10**	**P20**	**CR10**	**CR20**	**MAP**	**Rel_Ret**	**F-measure**
Visual	0.1200	0.1160	0.3357	0.4757	0.0095	416	0.1768
Text	**0.7680**	**0.7940**	0.5958	0.6358	**0.5164**	10370	**0.6710**
With PRF	0.6240	0.6640	**0.7106**	0.8192	0.4234	9216	0.6645
Combined	0.5680	0.6200	0.6902	**0.8242**	0.4585	**9594**	0.6641

We note a significant difference between the precision and cluster recall at ten (P10 & CR10) and at 20 (P20 & CR20) retrieved results. Unexpectedly, precision increases with retrieved results and up to the top hundred results. This could be due to some noisy early results introduced by the errors in visual retrieval. Contrary to ImageCLEFPhoto 2008 the F-measure was computed this year using a cut-off of the first ten results, which was a disadvantage to our method. The MAP and the Relevant Retrieved figures are promising and show consistency over the different topics.

Figure 2 shows the individual queries MAP performance of each of the four runs, while figure 3 shows the Cluster Recall at 10 retrieved results of the three textual and mixed runs. We note that the text-only run shows a higher standard deviation than the pseudo-relevance feedback method, especially due to the very low precision of two queries (Queries 10 and 43). In both cases the PRF method managed to reasonably answer the queries due to the visual input. We also remark that combining the cluster information in one query improves precision but decreases cluster recall.

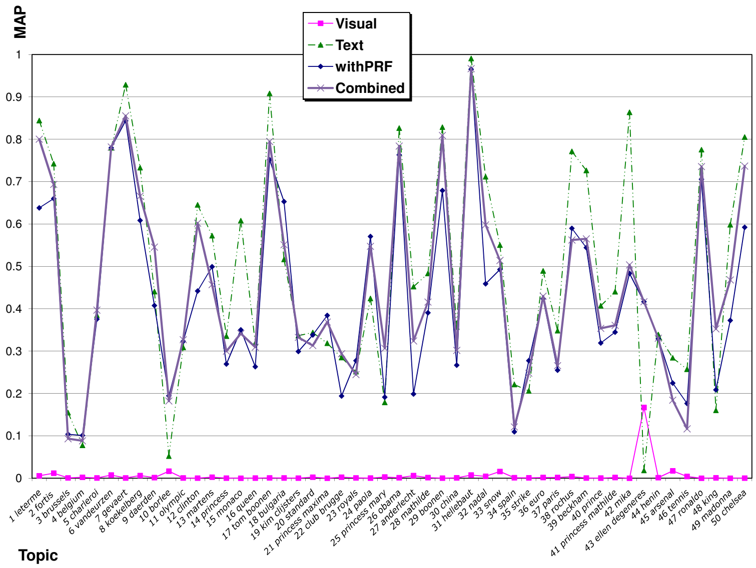

Fig. 2. Map by Query.

6 Conclusion

We experimented at ImageCLEF 2009 with applying semantic selectional restrictions to enhance cross-media pseudo-relevance feedback and different methods of query formulation for clustered queries. Our findings show that in the presence of valid results from a visual retrieval system, pseudo-relevance feedback can be successfully implemented and enhances the diversity of the results, except when the clusters are pre-determined. While the pre-querying semantic filtering applied in our approach can be useful, we will attempt combining it with a second post-retrieval filtering to remove noise and confirm the relevance of the results.

References

1. Paramita, M., Sanderson, M., Clough, P.: Diversity in Photo Retrieval: Overview of the ImageCLEFPhoto Task 2009. In: CLEF working notes 2009, Corfu, Greece. (2009)
2. El Demerdash, O., Kosseim, L., Bergler, S.: Image retrieval by inter-media fusion and pseudo-relevance feedback. In Peters, C., Deselaers, T., Ferro, N., Gonzalo, J., Jones, G.J.F., Kurimo, M., Mandl, T., Peas, A., Petras, V., eds.: CLEF. Volume 5706 of Lecture Notes in Computer Science., Springer (2008) 605–611
3. Stricker, M.A., Orengo, M.: Similarity of color images. In: Storage and Retrieval for Image and Video Databases (SPIE). (1995) 381–392
4. Mandal, M., Aboulnasr, T., Panchanathan, S.: Image indexing using moments and wavelets. IEEE Transactions on Consumer Electronics **42** (1996) 557–565

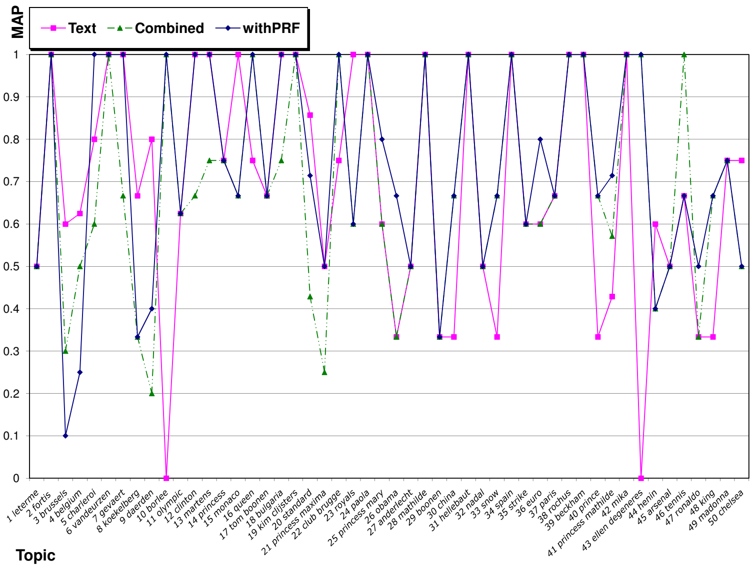

Fig. 3. Cluster Recall by Query.

5. Porter, M.F.: Snowball: A language for stemming algorithms. Published online (October 2001) Accessed 16.04.2009, 18.00h.
6. Ounis, I., Amati, G., Plachouras, V., He, B., Macdonald, C., Lioma, C.: Terrier: A High Performance and Scalable Information Retrieval Platform. In: Proceedings of ACM SIGIR'06 Workshop on Open Source Information Retrieval (OSIR 2006). (2006)
7. Fellbaum, C., ed.: WordNet An Electronic Lexical Database. The MIT Press, Cambridge, MA ; London (May 1998)
8. El Demerdash, O., Bergler, S., Kosseim, L.: CLaC at imageclef 2009. In: CLEF working notes 2009, Corfu, Greece. (2009)