Workshop of the Cross-Language Evaluation Forum for European Languages
CLEF 2007: Advances in Multilingual and Multimodal Information Retrieval pp 562-568 | Cite as

Text-Based Clustering of the ImageCLEFphoto Collection for Augmenting the Retrieved Results

Authors Authors and affiliations

Osama El Demerdash, Leila Kosseim, Sabine Bergler

Part of the Lecture Notes in Computer Science book series (LNCS, volume 5152)

Abstract

We present our participation in the 2007 ImageCLEF photographic ad-hoc retrieval task. Our

Text-Based Clustering of the ImageCLEFphoto Collection for Augmenting the Retrieved Results

Osama El Demerdash, Leila Kosseim, and Sabine Bergler

CLaC Laboratory - Department of Computer Science & Software Engineering - Concordia University
{osama_el,kosseim,bergler}@cse.concordia.ca

Abstract. We present our participation in the 2007 ImageCLEF photographic ad-hoc retrieval task. Our first participation in this year's ImageCLEF comprised six runs. The main purpose of three of these runs was to evaluate the text and visual retrieval tools as well as their combination in the context of the given task. The other purpose of our participation was to experiment with applying clustering techniques to this task, which has not been done frequently in previous editions of the ImageCLEF Ad hoc task. We used the preclustered collection to augment the search results of the retrieval engines. For retrieval, we used two publicly available libraries; *Apache Lucene* for text and *LIRE* for visual retrieval. The clustered-augmented results reduced slightly the precision of the initial runs. While the aspired results have not yet been achieved, we note that the task is useful in assessing the validity of the clusters.

1 Introduction

In this paper, we present our participation in the 2007 ImageCLEF photographic ad-hoc retrieval task. The task deals with answering 60 queries of variable complexity from a repository of 20,000 photographic images in the IAPR TC-12 collection. A full description of the task and the collection can be found in [1]. Our first participation in this year's ImageCLEF comprised six runs. The main purpose of three of these runs was to evaluate the text and content-based retrieval tools in the context of the given task. We therefore would like to stress that the evaluation of these tools can only be considered under the given parameters of the task, including the queries, the image collection and our utilization of these tools.

The other purpose of our participation was to experiment with applying clustering techniques to this task, which has not been done frequently in previous editions of the ImageCLEF Ad hoc retrieval task. While this task of ImageCLEF was not intended for the evaluation of interactive methods, it could still be useful in the evaluation of certain aspects of such methods such as the validity of the initial clusters in our case.

2 Related Work

Clustering, as an unsupervised machine learning mechanism, has rarely been investigated within the context of the ImageCLEF ad-hoc Retrieval task. This

C. Peters et al. (Eds.): CLEF 2007, LNCS 5152, pp. 562–568, 2008.

could be due to that clustering methods lend themselves more readily to interactive tasks and iterative retrieval. In the IR field, clustering has been experimented with extensively [2]. Its different applications involve clustering the whole data collection, part of it or clustering only the search results. In [3], images are clustered using labels from the surrounding HTML text. [4] applied clustering to content-based image retrieval using the Normalized Cut (NCut) algorithm under a graph representation. Another spectral clustering algorithm, *Locality Preserving Clustering (LPC)*, was introduced in [5] and found to be more efficient than NCut for image data. There is very little work in the literature on clustering using both content-based and text-based features. [6] and [7] describe successive clustering applied on text features then image features. The textual features comprised a vision based text segment as well as the link information while the *Color Texture Moments (CTM)*, a combined representation of color and texture were chosen for visual features. The only research we came across in the literature combining simultaneously image and textual features were from Microsoft Research Asia in 2005. [8] and [9] both use co-clustering techniques.

3 Resources

For retrieval, we used two publicly available libraries; *Apache Lucene* [10] for text and *LIRE* [11] for visual retrieval. Since our runs involved only English/English and Visual queries, we did not make use of any translation tools.

3.1 Text Retrieval

For text retrieval, we used the Apache Lucene engine, which implements a TF-IDF paradigm. Stop-words were removed, and the data was indexed as *field data* retaining only the *title*, *notes* and *location* fields, all of which were concatenated into one field. This helped reduce the size of the index, since our initial plan was to base the clustering on word-document cooccurrence and document-document similarity matrices. The number of indexed terms was 7577 from the 20,000 English source documents. All text query terms were joined using the *OR* operator. We did not apply any further processing of text queries.

3.2 Content-Based Retrieval

For visual retrieval, we employed v0.4 of the LIRE library which is part of the Emir/Caliph project available under the GNU GPL license. At the time of carrying out the experiments, LIRE offered three indexing options from the MPEG-7 descriptors: ScalableColor, ColorLayout and EdgeHistogram (a fourth one, Auto Color Correlogram, has since been implemented). The first two of these are color descriptors while the last is a texture one. We used all three indices. The details of these descriptors can be found in [12]. Only the best 20 images of each visual query were used. The visual queries consisted of the three images provided as example results. Thus, a maximum of 60 image results from visual queries were used in the evaluation.

4 Clustering Methodology

Three of our runs utilized preclustering of the data collection to augment the result set of the retrieval engines. Although we had intended in the beginning to cluster the results obtained from the text retrieval and content-based retrieval, we resorted to clustering the collection, given the small number of relevant results per query (compared to results from searching the World Wide Web for example).

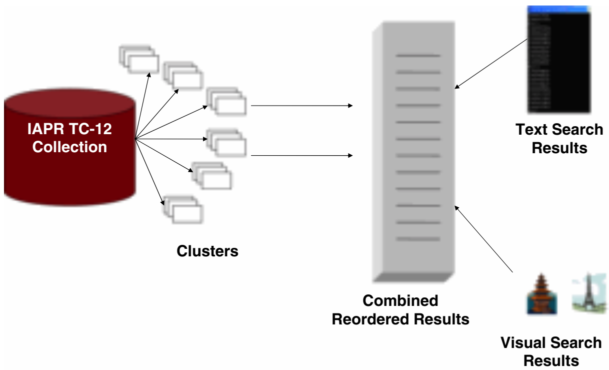

Fig. 1. Overview of the System

We employed a simple one-pass clustering algorithm which relied on forming clusters of the terms in the documents as they were processed. If a document's similarity to a cluster exceeded a certain threshold *(n)*, this document and its new terms were added to the term/document cluster. When a document was not associated with any cluster, it was temporarily assigned its own, which was deleted in the end if no other documents were associated with it. Also clusters larger than size *(s)* or smaller than size *(m)* were discarded since they were deemed unmeaningful. We did not, however, experiment with the parameters s and n and chose them with the little intuition we had about the data. The resulting clusters overlapped and did not cover all documents.

The following parameters were used in the experiments:

- Number of top results used for cluster expansion $t = 20$
- Number of results retained from image search $r = 20$
- Minimum number of common words to be in the same cluster $n = 3$
- Minimum size of cluster $m = 3$
- Maximum size of cluster $s = 300$

Figure 1 shows an overview of the system. In the mixed run (clacTXCB), we combined the results from Lucene text search and LIRE visual search by ranking the common ones highest, followed by all other text results and finally the visual results are added at the bottom of the list. This is due to the higher confidence we had in the text search results.

For augmenting the results from the clusters, we searched the clusters for the top t results and whenever one was found we inserted the other members of the cluster at this position in the result set, taking care not to include duplicate results from different clusters.

5 Results and Analysis

We submitted six runs at ImageCLEF 2007:

- clacTX: uses Lucene for text search
- clacCB: uses LIRE for visual search
- clacTXCB: combines the results from Lucene and LIRE
- clacCLSTX: augments clacTX with clusters
- clacCLSTCB: augments clacCB with clusters
- clacCLSTXCB: augments clacTXCB with clusters.

Table 1 shows the results our runs obtained at ImageCLEFphoto 2007 as well as the average, median and best results of the track. Our highest ranked run, (clacTXCB), is the one that combined results from Lucene (text retrieval) and LIRE (visual retrieval), getting a higher MAP as well as better performance on all other measures than the other runs. For this run, we used a combined list of the results from both engines, ranking common results highest on the list as described in Section 4.

The poor performance of our text-only run (clacTX) can be mainly attributed to the absence of stemming and query expansion/feedback. The total number of terms in the text index is 7577. When using a stemmer this figure is reduced by approximately 800. The results improve by an order of 1% to 2%.

As for query expansion, we estimate that the results can improve significantly by employing geographical gazetteers as well as synonyms. Indeed, further examination of the results shows that our poorest results were obtained for queries that reflect a combination of these two factors. For example, our poorest precision was obtained for topics no. 40 *(tourist destinations in bad weather)* and 41 *(winter landscape in South America)*.

Table 1. Results at ImageCLEF 2007

Run ID	Modality	MAP	P10	P20	P30	GMAP	Rel
clacTXCB	Mixed	0.1667	0.2750	0.2333	0.1599	0.0461	1763
clacCLSTXCB	Mixed	0.1520	0.2550	0.2158	0.1445	0.0397	1763
clacTX	Text	0.1355	0.2017	0.1642	0.1231	0.0109	1555
clacCLSTX	Text	0.1334	0.1900	0.1575	0.1205	0.0102	1556
clacCB	Visual	0.0298	0.1000	0.1000	0.0584	0.0058	368
clacCLSTCB	Mixed	0.0232	0.0817	0.0758	0.0445	0.0038	386
Average run	N/A	0.1292	0.2262	0.1913	0.1724	0.0354	1454
Median run	N/A	0.1327	0.2017	0.1783	0.1659	0.0302	1523
Best run	Mixed	0.3175	0.5900	0.4592	0.3839	0.1615	2251

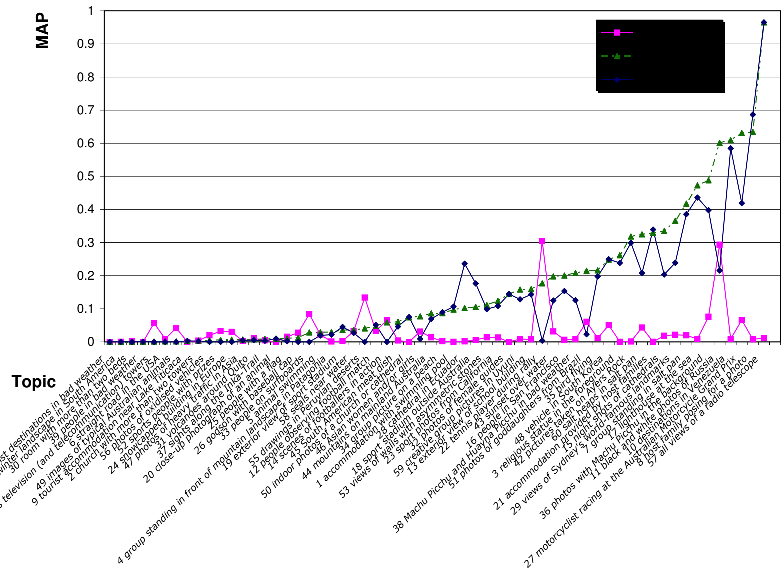

Fig. 2. Comparison between Text, Visual and Combined Results by Topic

Figure 2 shows the detailed performance by topic of the visual and text systems as well as their combination. We deduce from this figure that the run combining both systems had better results than those using the text or the visual system alone in a majority of the topics. In the few cases where the text retrieval obtained a higher MAP, the combined result was affected by noise induced from the visual results. On the other hand, the visual results achieved higher precision in some topics because of our reliance mainly on the text results, due to the higher confidence we had in them.

Our simple method of augmenting results using the preclustered data deteriorated the results in all three cases: text, visual and their combination. The main reason is that our clusters were less fine-grained than the requirements of the queries. We retained only 84 clusters of which only a handful were useful. When we experimented with the parameters we found that basing the clustering on a higher number of common words would lead to improving the results over the runs that do not employ the clusters. The one-pass clustering algorithm was unable to find this optimal parameter.

As for the other parameters described in section 4, they did not count for significant changes in the results. The number of results retained per visual query (=20) was found to be the most appropriate. Increasing or decreasing it degrades the precision. The same observation applies to the number of top documents (=20) used in augmenting the results, which can be attributed to the degrading precision after the top 20 as can be seen from the results. For the size of clusters we noted that very small clusters, which number below 30, were not useful since it is rare that one of their members happens to be in the query results. On the other hand, large clusters (with size > 200) introduce noise and reduce the precision.

6 Conclusion and Future Work

We intend to experiment with clustering the result set as well as introducing query expansion and pseudo-relevance feedback. Our final target is clustering based on both text and visual features.

Our first participation at ImageCLEF was satisfactory in that we were able to evaluate the IR tools we chose, as well as the validity of the initial clusters produced from a simple unsupervised clustering method.

References

1. Grubinger, M., Clough, P., Hanbury, A., Müller, H.: Overview of the ImageCLEF-photo 2007 Photographic Retrieval Task. In: Working Notes of the 2007 CLEF Workshop, Budapest, Hungary (September 2007)
2. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008), Online 17/08/2007, `http://www-csli.stanford.edu/schuetze/information-retrieval-book.html`
3. Sunayama, W., Nagata, A., Yachida, M.: Image Clustering System on WWW Using Web Texts. In: Proceedings of the Fourth International Conference on Hybrid Intelligent Systems (HIS 2004), pp. 230–235 (2004)
4. Chen, Y., Wang, J.Z., Krovetz, R.: Content-based Image Retrieval by Clustering. In: MIR 2003: Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval, pp. 193–200. ACM Press, New York (2003)
5. Zheng, X., Cai, D., He, X., Ma, W.Y., Lin, X.: Locality Preserving Clustering for Image Database. In: MULTIMEDIA 2004: Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 885–891. ACM Press, New York (2004)
6. Cai, D., He, X., Ma, W.Y., Wen, J.R., Zhang, H.: Organizing WWW Images Based on the Analysis of Page Layout and Web Link Structure. In: Proceedings of the 2004 IEEE International Conference on Multimedia and Expo, ICME 2004, Taipei, Taiwan, 27-30 June 2004, pp. 113–116. IEEE, Los Alamitos (2004)
7. Cai, D., He, X., Li, Z., Ma, W.Y., Wen, J.R.: Hierarchical Clustering of WWW Image Search Results Using Visual, Textual and Link Information. In: MULTIMEDIA 2004: Proceedings of the 12th annual ACM International Conference on Multimedia, pp. 952–959. ACM Press, New York (2004)
8. Li, Z., Xu, G., Li, M., Ma, W.Y., Zhang, H.J.: Grouping WWW Image Search Results by Novel Inhomogeneous Clustering Method. In: Chen, Y.P.P. (ed.) 11th International Conference on Multi Media Modeling (MMM 2005), pp. 255–261. IEEE Computer Society, Los Alamitos (2005)
9. Gao, B., Liu, T.Y., Qin, T., Zheng, X., Cheng, Q.S., Ma, W.Y.: Web Image Clustering by Consistent Utilization of Visual Features and Surrounding Texts. In: MULTIMEDIA 2005: Proceedings of the 13th Annual ACM International Conference on Multimedia, pp. 112–121. ACM Press, New York (2005)

10. Gospodnetic', O., Hatcher, E.: Lucene in Action (2005)
11. Lux, M., Granitzer, M.: Retrieval of MPEG-7 Based Semantic Descriptions. In: BTW-Workshop WebDB Meets IR at the GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web, University Karlsruhe (2005)
12. J.M. (ed.): MPEG-7 Overview (version 10). Technical Report N6828, ISO/IEC JTC1/SC29/WG11 (MPEG) (October 2004), online August 17, 2007, http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm