I2R AT IMAGECLEF PHOTO RETRIEVAL 2009

Sheng GAO and Joo-Hwee LIM
Institute for Infocomm Research, A*STAR, Singapore
1 Fusionopolis Way, Singapore, 138632

Abstract

In the paper, we introduce our systems and methods to promote the diversity of the ad-hoc photog retrieval in ImageCLEF 2009. The image database in this year is quite different from the previous years, not only increasing the corpus size from 20,000 images to half millionm but also changing the domain from the travel to news. Most of queries are related to person names and the text information in image documents is rich. Thus we put a lot of effort on text while using the visual information to elimiate or down-rank similar images in order to make top images visually dissimilar. To reduce the ambiguarity of query and infer the implicit dimension of diversity space, name entity extraction is applied on the top documents in order to extract informative phrase patterns to faciliate query expansion. Name entity based query expansion makes our system placed in the top performance.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; H.2.3 [Database Management]: Languages—Query Languages

General Terms

Measurement, Performance, Experimentation

Keywords

Language model, information retrieval, re-ranking, name entity extraction

1. Introduction

In the past years, the task of the photographic image retrieval works on the images in the travel domain and the database only has 20,000 images. The image document includes two modalities i.e. visual image and text description. Thus the information needs of the user query are formulated as keywords or image exemplars. In this year, the domain of the benchmark image database is changed to the news. The image content (visual and text description) in the news domain is quite different from the previous. So do the information needs, i.e. query types. In news domain, most of the queries and text descriptions of images are related with the personal names and events. Here are two query examples, one having the description about the meaning of diversity in the field of <*clusterTitle*> besides the key information need in the field of <*title*> while another only has key information frequently having one keyword.

Example 1
<top>
<num> 1 </num>
<title> leterme </title>
<clusterTitle> yves leterme </clusterTitle>
<clusterDesc> Relevant images contain photographs of Yves Leterme. Images of Leterme with other people are relevant if Leterme is shown in the foreground. Images of Leterme in the background are irrelevant. </clusterDesc>
<image> belga28/05980958.jpg </image>
<clusterTitle> leterme albert </clusterTitle>
<clusterDesc> Relevant images contain photographs of Yves Leterme and King Albert II. Images with only one of them are considered to be irrelevant. </clusterDesc>
<image> belga27/05960161.jpg </image>
<clusterTitle> leterme -albert </clusterTitle>
<clusterDesc> Images which contain photographs of Leterme which are not part of the above categories are relevant to this cluster. </clusterDesc>
<image> belga32/06229323.jpg </image>
</top>

Example 2
<top>
<num> 26 </num>

<title> obama </title>
<image> belga30/06098170.jpg </image>
<image> belga28/06019914.jpg </image>
<image> belga30/06107499.jpg </image>
</top>

The above two queries are all about the personal names. The key information need in the query only has one keyword, i.e. *leterme* and *obama*. Obviously they have great ambiguity because there are many personal names containing the *leterme* or *obama*. The diversity search prefers the system to return as much as both relevant and diverse images in the top document to reflect the ambiguity of the query. This is the promotion of photo retrieval task this year and last year.

In the two-modality based image retrieval, the diversity is characterized by the image visual content or the text descriptions attached to the image. Thus we can improve the diversity performance by grouping the visually-similar images and only selecting one or a few representative images in a group to represent the whole cluster. The visual similarity can be measured by matching one against another to see whether they are significantly different [2]. We can also find cues of similarity among image documents from analyzing text descriptions of image documents. The text cues may come from the bag-of-words, name entities, etc.

From our past experiences of participating photo retrieval [3, 4], the best system efficiently combine the text-based ranking system with the visual-based system. Thus our system architecture in this year also uses the two modalities. Simple and computation efficient visual features such as global colour moments and histogram, which work well in the previous data, do not operate well in this new dataset, because the content of images in this year are objects such as person. It requires us to extract other effective visual features, e.g. histogram of orientation which achieves higher performance in person detection [7]. However, computation cost of visual feature extraction and selection is very high for the large-scale database in comparison with the text feature. Thus, we pay much attention on text feature in this year and reuse the tools developed in the past for visual feature extraction.

2. System description

As introduced in the above, the text-based ranking system and the content-based image ranking system are individually built based on the text descriptions and visual descriptors respectively. To down-rank the visually similar images, we also apply the visual descriptors to re-rank the output from the text-based system. Before we discuss the details of submitted runs, we first discuss the index structure for large-scale retrieval system. This year the database has about half million image documents while it only has ~20,000 images in the previous corpus. It is crucial to build an index structure for efficiently and effectively processing the query rather than linearly scanning the database. For the content-based image retrieval system, the visual content of each image is characterized by a high dimensional feature. Although there are a lot of methods to index the documents, we apply the multi-probe locality sensitive hashing (MPLSH) [1][1], which is proved to be superior to the original LSH [8]. For the text-based retrieval system, we have shown the success of language model based information retrieval in the past years [3, 4]. Thus we still use the approach to index the text documents and process the query. For efficiency issue, we use the language model based search engine toolkit, *LEMUR*, developed by UMASS and CMU in this year[2]. Using these two toolkits, the query response of the systems is fast, which significantly reduce the time of tuning systems.

In the next, we will introduce five runs submitted for the official evaluation.

Run 1: LRI2R_TCT_TXT

It is the basic run submitted in order to validate the efficiency of name entity extraction for the query expansion. In the run, only the text modality is used. The keywords of the query are the words occurred in the fields of title and cluster title without using the cluster description. Each field represents a different information need of the user. For example, the title is a general information need while the cluster title gives a detailed information need. Thus, we formulate multiple sub-queries using these filed and then linearly combine multiple rank lists to reach a final rank list for the query. For example, for the first query discussed in Section 1, we have 4 sub-queries and thus the final rank is from fusing four rank lists. Without any prior knowledge of which field should be weighed more, the equal weight is used.

Run 2: LRI2R_TI_TXT

The run is to evaluate query expansion using name entity extraction. The name entity extraction operates on the text descriptions corresponding to the top-N (In the experiment, N=20) image documents in the Run 1. The name entity extraction tags the words in the text description using four categories: person, organization, location and others. We will only consider the first two categories. Then we search the phrase patterns occurred closely with the query keywords occurred in the title and cluster title used in Run 1, in a window size of 10 words. We assume that each phrase pattern

[1] http://lshkit.sourceforge.net/
[2] http://www.lemurproject.org/

may represent one dimension in the diversity space and we treat each phrase pattern as an individual sub-query similar to Run 1. Thus, we have mined multiple cluster titles automatically for each query. Then the final ranking is gotten using the same operation in Run 1.

Here we show the found cluster titles for the two queries in the above:

Example 1 (query 1)
leterme Yves
leterme Belga Photo Mark Renders
leterme Elio Di

Example 2 (query 26)
obama Barack
obama Senate
obama Michelle

Run 3: LRI2R_FUSE_TCTI_TXT

The run is just to combine the above two runs linearly and equally.

Run 4: LRI2R_DIVERSITY_TCTI_TXTIMG

This run is based on the Run 3. We use the visual feature to re-rank the rank list output from the Run 3. The visual feature we used is derived from the SIFT descriptor using the bag-of-visterm. In our implementation, two visual dictionaries with the size of 512 and 256 are individually generated using k-means on 20,000 randomly selected images from the database. We use the heuristic way similar to [9] to re-rank initial list. The procedure is shown in Figure 1. Thus we not only keep the order of the initial rank but also move the visually similar images into the bottom of the rank. That makes the top images visually dissimilar as much as possible.

```
Input: Initial rank
Output: Diversity rank
 1. Empty the diversity stack and non-novelty stack
 2. Push the top-1 in initial rank into diversity stack and remove it from the latter stack.
 3. Do
      a. Choose next document in initial rank and calculate its visual similarity with the documents in diversity stack
      b. If maximal similarity is lower than a threshold, then the document is novel, push it into diversity stack.
         Otherwise, push it into non-novelty stack
      c. Remove the document from the initial rank
 4. While (initial rank is not empty)
 5. Put all document in non-novelty stack in the diversity stack.
```

Figure 1 Diversity rank using visual feature

Run 5: LRI2R_TCTI_TXTIMG

The fun is a linear combination run between Run 3 and a run of content based image retrieval system. The CBIR run is similar to our past system in [3, 4], which uses three types of visual features, i.e. pyramid histogram of oriented gradients, Gabor texture and HSV-space histogram. From our initial analysis on the rank list of CBIR, its performance is bad and there are few images relevant in the top-10 besides.

In the 5 runs, only the last run uses the query image exemplars while the others mainly depend on the text analysis except for the run 4 uses the visual feature for re-ranking. There is only the last run is two-modality based retrieval while the others are the text-based retrieval.

3. Results

We list the official evaluation results in Tables 1-3. The overall performance on 50 queries is shown in Table 1. We can see the best run is the Run 2, which uses name entity extraction for query expansion. Comparing with its baseline, i.e. Run 1, its precision at the top-10, P@10, has a significant improvement from 0.79 to 0.848, while the cluster recall, CR@10 is increased to 0.671 from 0.657. This makes the best run achieves F-measure 0.7492. The achievement name entity extraction can mine the useful pattern for disambiguate the query and to discover the other dimensions of diversity space which are not in the query. The discovered diversity dimensions play an even important role for the queries in part 2, where the query only has a few words, most only having one word and without any diversity indicator, i.e. cluster title. This observation can be enhanced when analysing the performance on the queries in part 2 in Table 3. For this type of query, the best run is Run 2 while the Run 1 becomes the worst. Their gap in terms of diversity metric

(CR@10, P@10, F-measure) is even bigger. The F-measure in Run 2 is 0.7528 compared with 0.6556 in Run 1. However, when we analyze their diversity performance in the queries of part 1 in Table 2, it is seen that the Run 1 is better than the Run 2. It indicates that the manually selected cluster titles are superior to our automatically selected. This is reasonable because the automatically found phrase patterns may have a few noises. But the performance gap in this case only has ~3% in F-measure, not such bigger as in Table 3, where the F-measure difference is ~9%. The analysis strongly demonstrates name entity extraction is a very useful technology for diversity search, especially when the query is ambiguous and diversity dimensions are not indicated.

Now we analyze how the visual based re-ranking effects on the diversity by comparing the diversity performance between Run 4 and its corresponding baseline, Run 3. From Table 1, the F-measure for Run 4 is 0.6987 compared with 0.6842 for the Run 3 on 50 queries. Thus, visual re-ranking has a little benefit on the ranking performance. Its effect on the two types of queries is similar when analyzing their individual metrics in Table 2 and 3.

The linear fusion is widely used when combining multiple rank lists. What does it work on diversity search? The Run 3 is a combination of Run 1 and Run 2. From Table 1, we find that the Run 3 is the worst of all 5 runs, although individually the Run 1 and Run 2 are in the second and first rank position. The Run 5 is a combination of Run 3 and a CBIR system, which has a little improvement over Run 3. When combining the CBIR, the improvement is not such significant as our past years' systems [3, 4]. The reason should be caused by the poor performance of CBIR this year. Our experiences on CBIR in the travel domain cannot work well in the news domain. Thus, more powerful visual features for object images must be developed and person identification technology should work in the domain.

Rank	System Name	Query Type				Modality		CR@10	P@10	map	F-measure
		Title	Cluster Title	Cluster Desc	Image	Text	Image				
6	LRI2R_TI_TXT	v			v	v		0.6710	0.848	0.429	**0.7492**
13	LRI2R_TCT_TXT	v	v			v		0.6570	0.79	0.5033	**0.7174**
19	LRI2R_DIVERSITY_TCTI_TXTIMG	v	v		v	v	v	0.6239	0.794	0.4219	**0.6987**
22	LRI2R_TCTI_TXTIMG	v	v		v	v	v	0.6005	0.804	0.4779	**0.6875**
25	LRI2R_FUSE_TCTI_TXT	v	v		v	v		0.5965	0.802	0.4778	**0.6842**

Table 1 Ranking performance on 50 queries

Rank	System Name	Query Type				Modality		CR@10	P@10	map	F-measure
		Title	Cluster Title	Cluster Desc	Image	Text	Image				
8	LRI2R_TCT_TXT	v	v			v		0.7329	0.828	0.4904	**0.7760**
17	LRI2R_TI_TXT	v			v	v		0.6519	0.868	0.4324	**0.7446**
23	LRI2R_TCTI_TXTIMG	v	v		v	v	v	0.6055	0.852	0.5129	**0.7079**
26	LRI2R_DIVERSITY_TCTI_TXTIMG	v	v		v	v	v	0.6082	0.832	0.4445	**0.7027**
27	LRI2R_FUSE_TCTI_TXT	v	v		v	v		0.5975	0.848	0.5127	**0.7010**

Table 2 Ranking performance on 25 queries in query part 1

Rank	System Name	Query Type				Modality		CR@10	P@10	map	F-measure
		Title	Cluster Title	Cluster Desc	Image	Text	Image				
3	LRI2R_TI_TXT	v			v	v		0.6901	0.818	0.426	**0.7528**
11	LRI2R_DIVERSITY_TCTI_TXTIMG	v	v		v	v	v	0.6395	0.756	0.399	**0.6929**
23	LRI2R_FUSE_TCTI_TXT	v	v		v	v		0.5955	0.756	0.443	**0.6662**
24	LRI2R_TCTI_TXTIMG	v	v		v	v	v	0.5955	0.756	0.443	**0.6662**
31	LRI2R_TCT_TXT	v	v			v		0.5811	0.752	0.516	**0.6556**

Table 3 Ranking performance on 25 queries in query part 2

4. Conclusion

In the paper we describe the details of our submitted runs and analyze their diversity performance. Our analysis indicates two useful technology for promoting diversity search: 1) name entity extraction based query expansion and 2) visual based re-ranking. Name entity extraction can mine the phrase patterns to characterize the diversity space of information need of the user while the visual based re-ranking can further update ranking to make the top document visually dissimilar. Our successfull systems do not exploit the query image exemplars. The system of combining text-based retrieval and content-based image retrieval does not give an obvious gain. In future, we will dig into the two successful technologies to further improve the diversity performance.

References

1. W. Dong, Z. Wang, W. Josephson, M. Charikar & K. Li. Modeling LSH for performance tuning, Proc. of CIKM’08.
2. R. H. van Leuken, L. Garcia & X. Olivares, Visual diversification of image search results, Proc. of WWW’09.
3. S. Gao, J.-P. Chevallet & J.-H. Lim, IPAL at CLEF 2008: mixed-modality based image search, novelty based re-ranking and extended matching, Working Notes for the CLEF 2008 Cross Language Evaluation Forum.
4. S. Gao, J.-P. Chevallet, T. H. D. Le, T. T. Pham & J.-H. Lim, IPAL at ImageClef 2007 mixing features, models and knowledge, Working Notes for the CLEF 2007 Cross Language Evaluation Forum.
5. Jay M. Ponte & W. Bruce Croft, A language modeling approach to information retrieval, Proc. of SIGIR’98.
6. J. R. Finkel, T. Grenager & C. Manning, Incorporating non-local information into information extraction systems by Gibbs sampling, Proc. of ACL’05.
7. N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, Proc. of CVPR’05.
8. A. Andoni & P. Indyk, Near-optimal hashing algorithms for approximate nearest neighbor in high dimension, Proc. of FOCS’06.
9. R. H. van Leuken, L. Garcia, X. Olivares & R. V. Zwol, Visual diversification of image search results, Proc. of WWW’09.