Increasing Relevance and Diversity in Photo Retrieval by Result Fusion

Yih-Chen Chang and Hsin-Hsi Chen
Department of Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan
E-mail: ycchang@nlg.csie.ntu.edu.tw; hhchen@csie.ntu.edu.tw

Abstract

This paper considers the strategies of query expansion, relevance feedback and result fusion to increase both relevance and diversity in photo retrieval. In the text-based retrieval only experiments, the run with query expansion has better MAP and P20 than that without query expansion, and only has 0.85% decrease in CR20. Although relevance feedback run increases both MAP and P20, its CR20 decreases 10.18% compared with non-feedback run. It shows that relevance feedback brings in relevant but similar images, thus diversity may be decreased. The run with both query expansion and relevance feedback is the best in the four text-based runs. In the content-based retrieval only experiments, the run without feedback outperforms the run with feedback. The latter has 10.84%, 9.13%, and 20.46% performance decrease in MAP, P20, and CR20. In the fusion experiment, integrating text-based and content-based retrieval not only reports more relevant images, but also more diverse ones.

1. Introduction

In the photo retrieval task of ImageCLEF 2008, the focus is shifted from cross language image retrieval to promote diversity. Besides relevance, retrieving diverse items representing different subtopics is also concerned. How to balance the relevance and diversity is challenging. This paper studies the strategies of query expansion and relevance feedback in text-based and content-based retrieval, and shows how to merge the results of text and image queries to increase both relevance and diversity.

This paper is organized as follows. Sections 2, 3 and 4 present text-based retrieval, content-based retrieval and combination of both, respectively. Section 5 shows the runs submmited for formal evaluation in the photo retrieval task, and discsses the effects of different retrieval and fusion strategies.

2. Text-Based Retrieval

In text-based retrieval, we consider the strategies of query expansion and relevance feedback. Assume the text corpus T is composed of n terms, $t_1, t_2, \ldots, t_n$, and a query Q contains m query terms, $q_1, q_2, \ldots, q_m$. We expand Q in the following way.

(1) For each corpus term t_i and query term q_j, compute $P(q_j \mid t_i)=P(t_i,q_j)/ P(t_i)$.

(2) For each corpus term t_i, compute $OverlapNum(t_i,Q)$ defined below.

$$OverlapNum(t_i,Q) = cardinality\{q \mid q \in Q, P(q \mid t_i)>0\}$$

(3) For all $t_i \in T$, if $\sum_{j=1}^{m} P(q_j \mid t_i) \times \mathrm{OverlapNum}(t_i,Q) > thd$, then t_i will be added to new query Q'. In the experiments, thd is set to 1. In other words, the original query terms which also appear in the corpus will be added into Q'.

We adopt Lemur as our text IR system. The weighting function is BM25 with parameters (K1=1.2, B=0.75, K3=7). For relevance feedback, we select the top-10 terms of the highest BM25 scores from the top-5 retrieved documents, and add them to the query. The expanded terms have 1/2 weight of the original query terms.

3. Content-Based Retrieval

For each image g_i, we extract two kinds of features: $SizeFeature(g_i)$ and $ColorFeature(g_i)$. These two functions are defined below.

(1) $SizeFeature(g_i)$ $= 0$, if $height(g_i)>width(g_i)$

$= 1$, if $height(g_i)<=width(g_i)$

(2) $ColorFeature(g_i)$: divide g_i into 32×32 blocks, and extract their RGB values.

The similarity of two images, g_i and g_j, is computed as follows.

(1) Compute the color similarity of g_i and g_j based on their color features.
$ColorSimilar(g_i, g_j)$ = number of blocks in g_i and g_j, whose R, G and B value differences are not larger than 10.

(2) Compute the size similarity of g_i and g_j based on their size features.
If $SizeFeature(g_i)$ and $SizeFeature(g_j)$ is the same, then $SizeSimilar(g_i, g_j)=1.5$. Otherwise, $SizeSimilar(g_i, g_j)=1.0$.

(3) The similarity of g_i and g_j is in terms of $SizeSimilar$ and $ColorSimilar$:
$Similar(g_i, g_j)= SizeSimilar(g_i, g_j) \times ColorSimilar(g_i, g_j)$

4. Combining Text-based and Content-based Retrieval

In image retrieval, we compute the similarities of the query images and all the images in the data set and select the most similar image for media mapping (Chen and Chang, 2006). The corresponding text description of the reported image is regarded as a text query for further retrieval. The results of text-based and content-base retrieval are merged in the following way. We normalize the scores of the two result lists by the corresponding top-1 scores (Tsai, Wang, and Chen, 2008), i.e., the normalized scores will be within 0 and 1, and merge the lists with the same weights by their normalized scores.

5. Experiments and Discussion

We submit 7 runs shown below for the formal evaluation.

(1) NTU-EN-EN-AUTO-NOFB-TXT
This run is baseline. We employ Lemur for text-based retrieval without query expansion and relevance feedback.

(2) NTU-EN-EN-AUTO-FB-TXT
This run employs Lemur for text-based retrieval with relevance feedback.

(3) NTU-EN-EN-AUTO-QE-NOFB-TXT
This run employs Lemur for text-based retrieval with query expansion.

(4) NTU-EN-EN-AUTO-QE-FB-TXT
This run employs Lemur for text-based retrieval with query expansion and relevance feedback.

(5) NTU-IMG-EN-AUTO-NOFB-TXTIMG.
This run employs content-based retrieval first, then adopts media mapping to transform the image query to text query, and employs Lemur for text-based retrieval without relevance feedback.

(6) NTU-IMG-EN-AUTO-FB-TXTIMG
This run is similar to NTU-IMG-EN-AUTO-NOFB-TXTIMG except that relevance feedback is done.

(7) NTU-EN-EN-AUTO-QE-FB-TXTIMG
This run merges the results of NTU-IMG-EN-AUTO-FB-TXTIMG and NTU-EN-EN-AUTO-QE-FB-TXT.

The evaluation of the formal runs is based on mean average precision (MAP), precision at 20 (P20) and instance recall at rank 20 (CR20), which calculates the percentage of different clusters represented in the top 20. Table 1 lists the experimental results of employing text query only. The run with query expansion has better MAP and P20 than that without query expansion, and only has 0.85% decrease in CR20. Although relevance feedback increases both MAP and P20 in EN-EN-AUTO-FB-TXT run, its CR20 decreases 10.18% compared with EN-EN-AUTO-NOFB-TXT. It shows that relevance feedback brings in relevant but similar images, thus diversity may be decreased. The run with both query expansion and relevance feedback is better than the other three runs. Compared with baseline, it has 33.79%, 44.44%, and 0.27% increase in MAP, P20 and CR20, respectively.

Table 1. Comparisons of Runs Employing Text Query Only

Runs	Feedback	Expansion	MAP	P20	CR20
EN-EN-AUTO-NOFB-TXT	No	No	0.1790	0.2077	0.2602
EN-EN-AUTO-QE-NOFB-TXT	No	Yes	0.1967 (+9.88%)	0.2244 (+8.04%)	0.2580 (-0.85%)
EN-EN-AUTO-FB-TXT	Yes	No	0.2122 (+18.54%)	0.2692 (29.61%)	0.2337 (-10.18%)
EN-EN-AUTO-QE-FB-TXT	Yes	Yes	0.2395 (+33.79%)	0.3000 (+44.44%)	0.2609 (+0.27%)

Table 2 lists the experimental results of employing sample images. In the experiments, 3 example images are considered. The run without feedback outperforms the run with feedback. The latter has 10.84%, 9.13%, and 20.46% performance decrease in MAP, P20, and CR20. The possible reason of the drop in precision is the top-5 retrieved images for feedback may be very specific. That may introduce noises. Consider topic 43, *sunset over water*, as an example. The correct image should contain both *sunset* and *water*. The query without feedback is "***Sunset*** *at the* ***sea*** *the dark outlines of a mountain in the foreground the sun is rising over the* ***sea*** *behind it a light orange sky in the background peru*". In the top-5 retrieved images, only one contains both scenes, but all of them contain *sunset* scene. There are 34 relevant images in the result list before feedback, and only 25 relevant images after feedback. The MAP decreases from 0.1776 to 0.0535 after feedback. CR20 decreases more than MAP and P20. It shows pure relevance feedback is harmful to diversity.

Table 2. Comparisons of Runs Employing Image Query Only

Run	Feedback	MAP	P20	CR20
IMG-EN-AUTO-NOFB-TXTIMG	No	0.2103	0.3090	0.1779
IMG-EN-AUTO-FB-TXTIMG	Yes	0.1875 (-10.84%)	0.2808 (-9.13%)	0.1415 (-20.46%)

Table 3 compares the performance of employing text query only, image query only, and both. The fusion run, which achieves MAP 0.2809, P20 0.3769 and CR20 0.2763, is the best of our 7 submitted runs in the formal evaluation. It shows that integrating text-based and content-based retrieval not only reports more relevant images, but also more diverse ones.

Table 3. Comparisons of Runs Employing Text Query, Image Query and Both

Run	Feedback	MAP	P20	CR20
EN-EN-AUTO-QE-FB-TXT	Yes	0.2395	0.3000	0.2609
IMG-EN-AUTO-FB-TXTIMG	Yes	0.1875	0.2808	0.1415
EN-EN-AUTO-QE-FB-TXTIMG	Yes	0.2809	0.3769	0.2763

6. Conclusion

This paper considers query expansion, relevance feedback and result fusion to deal with relevance and diversity in image retrieval. Query expansion is useful to increase the precsion in text-based retrieval, but has a little negative effect on the diversity. Relevance feedback is harmful to diversity when this strategy is used independently or in single type of queries. Text-based and content-based retrievals have their own special capability, so that both relevance and diversity are improved.

References

Hsin-Hsi Chen and Yih-Chen Chang (2006). "Language Translation and Media Transformation in Cross-Language Image Retrieval." *Proceedings of 9th International Conference on Asian Digital Libraries*, November 27-30, 2006, Kyoto, Japan, Lecture Notes in Computer Science, 4312, 350-359.

Ming-Feng Tsai Yu-Ting Wang and Hsin-Hsi Chen (2008). "A Study of Learning a Merge Model for Multilingual Information Retrieval." *Proceedings of the 31st Annual International ACM SIGIR Conference*, 20-24 July 2008, Singapore, 195-202.