CEA LIST's Participation at MediaEval 2013 Retrieving Diverse Social Images Task

Adrian Popescu
CEA, LIST, Vision & Content Engineering Laboratory, 91190 Gif-sur-Yvette, France.
adrian.popescu@cea.fr

ABSTRACT

Clustering is by far the most popular diversification technique described in literature. Its aim is to group together images that are related following some similarity criterion. Here we aim to tackle the problem differently and explore a reranking-based techniques that increase diversity by considering the "informativeness" of each new image with respect to the set of images that were already selected. "Informativeness" is defined using social cues, such as user ID and date, visual cues extracted from the low-level representation of the image or multimedia cues that combine visual and textual processing. For some of the runs, we also exploit an initial k Nearest Neighbors (k-NN) inspired image reranking that is meant to reduce the amount of noise present in the result set.

1. INTRODUCTION

An efficient information retrieval system should be able to summarize search results so that it surfaces results that are both relevant and that are covering different aspects of a query. Relevance was more thoroughly studied than diversification and, even though a considerable amount of diversification literature exists, the topic remains a hot one. Usually, given a set of items to diversify, results clustering is exploited in order to propose a diversified representation of that set [4]. Our purpose at MediaEval 2013 Diverse Images [1] is to build on our previous work [3] and adapt it to social image search. We aim to replace clustering by a simpler method that is based on the "informativeness" (i.e. the amount of novelty brought by every new image). We first describe the different cues that we use to approximate "informativeness" and a k-NN inspired image reranking procedure that aims to reduce the amount of noise in the result set. Then we introduce the reranking procedure used for results diversification. Finally, we present the submitted runs and discuss the results obtained.

2. DIVERSIFICATION CUES
2.1 Social Cues

Social cues were already successfully exploited in POI image diversification [2]. The most straightforward diversification methods rely on the initial Flickr ranking and exploit simple cues such as user ID or user ID associated to the day when the photo was taken. The first cue aims to maximize

the number of unique users that contribute to the results set. The intuition behind its use is that different users will photograph different aspects of a POI. The second cue is a lighter version of the first and it assumes that if a user returns to a POI on a different day, she is likely to photograph another aspect of it.

2.2 Visual Cues

The visual content of the images is often used in clustering-based diversification techniques. Although they do not convey semantic information directly, visual features can be useful, especially for topics with a small semantic coverage, such as points of interest. Preliminary tests realized with the different features provided by the organizers showed that HOG outperforms the other features, although the differences were not very significant. Given these preliminary results, we decided to exploit HOG features in our runs.

2.3 Textual Cues

We tried to exploit the textual models provided with the dev set but no accuracy improvement compared to the Flickr ranking was observed. This negative result might be explained by the fact that the precision of the Flickr ranking is already high. Consequently, we did not perform any textual processing and simply exploited the text-based ranking provided by Flickr in our runs.

3. RERANKING FOR NOISE REDUCTION

The initial result set is noisy and we introduce a k-NN inspired approach that exploits social and visual cues to rerank results. We considered all the images of the POI as a positive set and built a negative set of the same size by sampling images of other POIs from the collection. Then we compared the HOG features of each image to all other images' features from positive and negative sets and retained the top 5 most similar results. We counted the number of different users that contributed to the top 5 neighbors and, then the number of positive exemples in the top 5 neighbors and the average distance to the first 5 positive neighbors. These cues were cascaded to rerank images and the top 70% images from the reranked list are retained for experiments that exploit this reranking technique.

4. RERANKING FOR DIVERSIFICATION

Given an initial list of results to diversify, the purpose of this reranking step is to surface different aspects of the topic in the top results. Hash tables are created to store the unique

Table 1: Run performances with three official metrics: CR - cluster recall, P - precision, F1 - harmonic mean of CR and P. All values are expressed after 10 results. The first three columns present results obtained with expert annotations and the last three columns results obtained with crowdsourcing (averages over the three workers).

	Expert annotations			Crowd sourcing		
Run	CR	P	F1	CR	P	F1
#1	0.37	0.75	0.48	0.74	0.78	0.73
#2	0.42	0.77	0.52	0.73	0.71	0.68
#3	0.36	0.76	0.48	0.75	0.77	0.73
#4	0.40	0.74	0.5	0.71	0.69	0.67

combinations of diversification cues. To diversify results, we start from the initial ranking, create a temporary structure to store the diversification and initialize the reranked list with the first image. We assess the images from the list and add them to the diversified list only if they satisfy a "informativeness" criterion. This criterion is defined using the diversification cues described in Section 2. When we reach the end of the list, we reinitialize the temporary structure and choose images that are not already in the diversified reranking. The process is repeated until all images are added to the diversified list of results.

5. RESULTS AND DISCUSSION

We submitted four different runs at this year's Diverse Social Images Task [1]. These runs produced by using different types of cues and their combinations on the same dataset. Our submissions are briefly described below: **RUN1** is based on the HOG visual feature provided by the organizers. We first apply the visual reranking procedure described in 3 to reduce the amount of noise in the initial results and retain. Then, we initialize the diversified list with the first image and then add new images by maximizing their average visual distance with respect to the images that are already in the diversified list. **RUN2** is based on the initial Flickr ranking and on the hash table of unique users described in Subsection 2.1. In each diversification round, new images are selected only if there is another image of the same user was not already chosen in that round. **RUN3** is similar to RUN1 with a difference concerning the reranking for noise reduction. This reranking is done through a linear combination of the ranks of the images in the initial Flickr results set and of the ranks of the images in the HOG-based reranking exploited for RUN1. Empirical tests on the dev set showed that the optimal combination of results is that which gives a weight of 0.3 to the Flickr ranking and 0.7 to the HOG-based reranking. **RUN4** is similar to RUN2 but it exploits the user-date hash table instead of the user hash in order to diversify results.

The results in Table 1 show the best results for the expert annotations were obtained with the simplest reranking approaches, that exploit only social cues. The user-based reranking (RUN2), which performs only a slight alteration of Flickr results by maximizing the number of different users represented in the top results, had the best performances. The assumption that different users will capture different aspects of a POI seems to be validated. The exploitation of

the user-date combination (RUN4) produces a performance loss compared to RUN2. The good CR@10 scores obtained for RUN2 and RUN4 indicate that the diversification technique based on social cues is efficient. The improvement of diversity is accompanied by a small improvement of P@10 for RUN2 and by a small precision loss for RUN4. Consequently, the F1@10 measure, which combines relevance and diversity is improved w.r.t. the original Flickr ranking. RUN1 and RUN3, which are based on the exploitation of visual and multimedia cues have performances that are inferior to those of RUN2 and RUN4. They rely on more complex processing, which includes the maximization of the visual diversity of results, but this processing does not seem to be useful for the test set.

When considering the crowd sourcing ground truth, the results obtained with social cues (RUN2, RUN4) are inferior to the results obtained with visual and multimedia processing (RUN1 and RUN3). However, the difference CR@10 between the best and the worst run is small and it is difficult to have definitive conclusions based on these scores.

6. CONCLUSIONS

The results obtained on the expert annotation of the test set are surprising since initial tests performed on the dev set gave the following performance order: RUN3, RUN1, RUN2 and RUN4. On the test set, only the order of RUN2 and RUN4 is respected. The results obtained on the crowd sourcing ground truth are more inline with those obtained on the development set. The run performances that we obtained during the campaign confirm the findings of [2] usefulness of social cues in result diversification. The small effect of visual cues is in contradiction with the results of [2] and [3] but we need to investigate further the reasons of these poor performances. One explanation might come from the poor adaptation of HOG, a simple global descriptor, to the application domain - i.e. tourism photos. In the future, we plan to explore the integration of social and visual cues in order to obtain a more efficient diversification.

7. ACKNOWLEDGMENT

8. REFERENCES

[1] B. Ionescu, M. Menendez, H. Muller, and A. Popescu. Retrieving diverse social images at mediaeval 2013: Objectives, dataset and evaluation. In *MediaEval 2013 Workshop, CEUR-WS.org, ISSN: 1613-0073*, Barcelona, Spain, October 18-19 2013.

[2] L. S. Kennedy and M. Naaman. Generating diverse and representative image search results for landmarks. In *Proc. of WWW 2008*, pages 297–306, New York, NY, USA, 2008. ACM.

[3] A. Popescu, P.-A. Moëllic, I. Kanellos, and R. Landais. Lightweight web image reranking. In *Proc. of ACM Multimedia 2009*, pages 657–660, New York, NY, USA, 2009. ACM.

[4] R. H. van Leuken, L. Garcia, X. Olivares, and R. van Zwol. Visual diversification of image search results. In *Proc. of WWW 2009*, pages 341–350, New York, NY, USA, 2009. ACM.