

Experiments in Diversifying Flickr Result Sets

Neha Jain
nj1g12@ecs.soton.ac.uk

Jonathon Hare
jsh2@ecs.soton.ac.uk

Sina Samangooei
ss@ecs.soton.ac.uk

John Preston
jlp1g11@ecs.soton.ac.uk

Jamie Davies
jagd1g11@ecs.soton.ac.uk

David Dupplaw
dpd@ecs.soton.ac.uk

Electronics and Computer Science, University of Southampton, United Kingdom

ABSTRACT

The 2013 MediaEval *Retrieving Diverse Social Images Task* looked to tackling the problem of search result diversification of Flickr results sets formed from queries about geographic places and landmarks. In this paper we describe our approach of using a min-max similarity diversifier coupled with pre-filters and a reranker. We also demonstrate a number of novel features for measuring similarity to use in the diversification step.

1. INTRODUCTION AND MOTIVATION

The diversification of search results is increasingly becoming an important topic in the area of information retrieval. The 2013 MediaEval *Retrieving Diverse Social Images Task* [4] aimed to foster new multimodal approaches to the diversification of result sets from social photo retrieval.

Our motivation for this task was to build on the diversification techniques we developed in ImageCLEF'09 [1] by incorporating truly multimodal data. We were also motivated to explore how the precision of the search results could be improved by filtering and re-ranking prior to the diversification step, thus minimising the loss in precision usually seen when diversification is applied.

2. METHODOLOGY

In terms of overall approach, after a number of experiments, we settled on the workflow illustrated in Figure 1. In order to improve precision, we applied filters to the input results list to remove images unlikely to be relevant, and for the runs that allowed use of the text and metadata, we reranked the results before applying diversification. To diversify the results, after testing a number of techniques (i.e. clustering followed by round-robin selection), we reverted to a Min-Max diversification technique as it gave the best results on the development dataset with the features we used.

Briefly, the Min-Max technique takes as input a similarity matrix and a pivot image, and uses this to build a result list. The pivot is taken as the first image in the result list. The second image is chosen as the one that has the minimum similarity to the pivot. The remaining images are chosen such that they have the maximum dissimilarity to all of the previously chosen images. Similarity of an image from a set of images can be computed via a number of functions

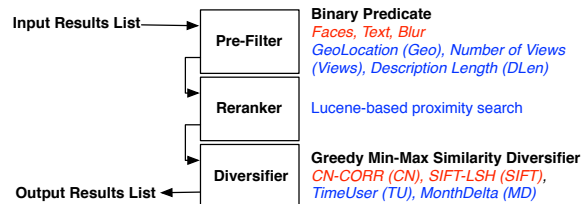


Figure 1: Overall workflow for diversify results

over the piecewise similarities of the image to each element of the set; *sum*, *product* and *max* are typical choices. On the development set, we found *max* worked best. The implementation of our methodology was realised in Java using OpenIMAJ¹ [2] and Lucene².

2.1 Pre-Filters

We removed many of the images which weren't relevant using some of our pre-filters before diversifying them. Images which contained frontal or side-views of faces in focus were discarded and we also got rid of the blurred out-of-focus ones. Images were further checked for the amount of text they contained and those with high percentage were thrown away. Those images which had been geotagged more than 8 km away from their actual location were removed. We found that images without any views were usually not relevant and so only took into consideration those which had more than 2 views. Similarly we discovered that images with very large descriptions tend to be irrelevant and hence filtered out those whose descriptions were over 2000 characters long.

2.2 Reranker

The original results lists provided in the task were retrieved by searching Flickr with a given monument name. The exact search implementation used by Flickr is unknown, but it is likely to be a variant of the vector space model with stemming. A better, more precise, ranking of the results can be achieved by performing a phrase or proximity search in which the results are scored higher if the query terms occur in close proximity in the metadata. To apply proximity-based reranking, we indexed the title, description and tags fields of each image in the filtered results list with Lucene, and performed the following query: (TITLE:"monument"~20 OR TAGS:"monument"~20 OR DESCRIPTION:"monument"~20)

Copyright is held by the author/owner(s).

MediaEval 2013 Workshop, October 18-19, 2013, Barcelona, Spain

¹<http://openimaj.org>

²<http://lucene.apache.org>

Table 1: Run configuration

(a) Pre-filters applied in each of the runs.

Run	Visual			Meta/Textual		
	Face	Blur	Text	Geo	Views	DLen
1	✓	✓	✓			
2				✓	✓	✓
3		✓	✓	✓	✓	✓

(b) Reranker and features in each of the runs.

Run	Reranker	Visual		Meta/Textual	
		CN	SIFT	TU	MD
1		✓			
2	✓			✓	✓
3	✓		✓	✓	✓

Table 2: Official Results. Crowdworker evaluation was performed on a subset of 50 locations from the testset (346 locations).

Run	Expert			Crowdworker							
	ALL			P@10	GT1		GT2		GT3		
	P@10	CR@10	F1@10		CR@10	F1@10	CR@10	F1@10	CR@10	F1@10	
1	0.6994	0.4081	0.4926	0.6612	0.8174	0.7043	0.7858	0.688	0.6398	0.6197	
2	0.8231	0.4306	0.5397	0.7694	0.8124	0.7689	0.7474	0.7276	0.6745	0.6944	
3	0.8158	0.4398	0.5455	0.7714	0.8184	0.7734	0.7486	0.7263	0.668	0.6906	

OR (TITLE:"monument")^0.5 (TAGS:"monument")^0.1

2.3 Similarity Matrices

In order to use the Min-Max diversifier, a similarity matrix is required. At the beginning of the task we spent some time analysing the data, and looking at features which could be sensibly used to compute similarity. One particular problem we noticed was that many of the images had the same description and tags, even though they were visually diverse. This means that standard techniques for diversification based on the text are unlikely to work well in many cases, and would in all likelihood end up being similar to just diversifying based on the users that took the photo. With this in mind, we started to explore other features that could work better.

Color Naming Histogram (CN). The provided CN histogram features [4], were used to create a similarity matrix by using correlation to measure the pairwise similarity.

Scale-Invariant Feature Transform (SIFT). SIFT features from the images were extracted and hashed using an LSH scheme [3]. A sparse binary similarity matrix was created from these, by setting a similarity of 1 to pairs of images in which there was a hash collision.

Time User (TU). Images taken by the same user within a short time period are likely to be similar. A similarity matrix was constructed with the following constraints: pairs of images taken a less than a minute apart had similarity 1; images more than 3.25 mins apart had 0 similarity. Between 1 and 3.25 minutes the similarity falls off logarithmically.

Month Delta (MD). Similar to the TU feature, images have increasing similarity with closer month of year.

3. EXPERIMENTS AND RESULTS

Three runs were submitted; their configuration with respect to the methodology and features described in Section 2 is illustrated in Tables 1a and 1b. Where multiple features were used, the similarity matrices were just averaged to create a single matrix. Two major points can be noted from the results. Firstly, using textual and visual features outperforms the use of either modality alone with our techniques. It is also clear that the reranking stage massively helps improve precision. Secondly, the high variability in results

across the experts and crowdworkers indicates that the task is actually rather subjective; it is particularly interesting that when compared against the crowdworker groundtruths our cluster recall scores are almost double, perhaps indicating that the experts tended to over-segment the result sets.

4. CONCLUSIONS

In this work we’ve explored different features for search result diversification, and also explored how relevance can be maximised by pre-filtering and re-ranking prior to the diversification step. The results indicate that our re-ranking step gives a good increase in precision. The combination of features from multiple modalities leads to a modest increase in diversity. In the future we intend to investigate whether automatically generated classifications from the visual features (indoor/outdoor, etc) can be leveraged to increase diversity.

5. ACKNOWLEDGMENTS

The described work was funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreements 270239 (ARCOMEM), and 287863 (TrendMiner).

6. ADDITIONAL AUTHORS

Additional author: Paul Lewis (phl@ecs.soton.ac.uk)

7. REFERENCES

- [1] J. Hare, D. Dupplaw, and P. Lewis. IAM@ImageCLEFphoto 2009: Experiments on Maximising Diversity using Image Features. In *CLEF 2009 Workshop*, pages 42–42, September 2009.
- [2] J. S. Hare, S. Samangoeei, and D. P. Dupplaw. OpenIMAJ and ImageTerrier: Java libraries and tools for scalable multimedia analysis and indexing of images. In *ACM MM’11*, pages 691–694. ACM, 2011.
- [3] J. S. Hare, S. Samangoeei, D. P. Dupplaw, and P. H. Lewis. Twitter’s visual pulse. In *ICMR’13*, pages 297–298, New York, NY, USA, 2013. ACM.
- [4] B. Ionescu, M. Menéndez, H. Müller, and A. Popescu. Retrieving diverse social images at mediaeval 2013: Objectives, dataset and evaluation. In *MediaEval 2013 Workshop*, Barcelona, Spain, October 18-19 2013.