Recod @ MediaEval 2014: Diverse Social Images Retrieval

Rodrigo T. Calumby[1,2], Vinícius P. Santana[2], Felipe S. Cordeiro[2],
Otávio A. B. Penatti[3], Lin T. Li[1], Giovani Chiachia[1], and Ricardo da S. Torres[1]

1 RECOD Lab, Institute of Computing, University of Campinas (UNICAMP), Campinas, SP – Brazil
2 Department of Exact Sciences, University of Feira de Santana (UEFS), Feira de Santana, BA – Brazil
3 SAMSUNG Research Institute Brazil, Campinas, SP – Brazil

[rtcalumby, vpsantana, fscordeiro]@ecomp.uefs.br, o.penatti@samsung.com, [lintzyli, chiachia, rtorres]@ic.unicamp.br

ABSTRACT

This paper presents the results of the first participation of our multi-institutional team in the *Retrieving Diverse Social Images Task* at MediaEval 2014. In this task we were required to develop a summarization and diversification approach for social photo retrieval. Our approach is based on irrelevant image filtering, image re-ranking, and diversity promotion by clustering. We have used visual and textual features, including image metadata and user credibility information.

1. INTRODUCTION

Promoting diversity is an effective approach for improving retrieval results and user search experience. For instance, it has been applied for tackling ambiguous or underspecified queries, or producing summaries. The *Retrieving Diverse Social Images Task* [2] combines such problems into a challenge on visual summarization for social photo retrieval in a tourism related context. This paper presents our first efforts on relevance improvement and diversity promotion using image visual features, metadata and user credibility information.

2. PROPOSED APPROACH

The proposed approach follows the general pipeline presented in Figure 1. At first, two filtering steps are conducted in order to reduce the amount of irrelevant images. Afterwards, re-ranking steps are applied for improving image rank positions according to two different relevance aspects. Finally, clustering is performed and followed by representative and diverse images selection. Specific combinations of the proposed steps were set for each submitted run (Section 3).

2.1 Filtering

In order to reduce the number of non-relevant images we adapted two filtering strategies: Geographic filtering and Face filtering. Eliminating non-relevant images allows higher effectiveness in terms of final relevance and boosts the diversification procedure. This is a consequence of fewer non-relevant items as candidates for the final diversified list.

The geographic filtering (GeoFilter) takes the reference lat/long of each location and then eliminates all images located farther than a given range. In this case, only geo-
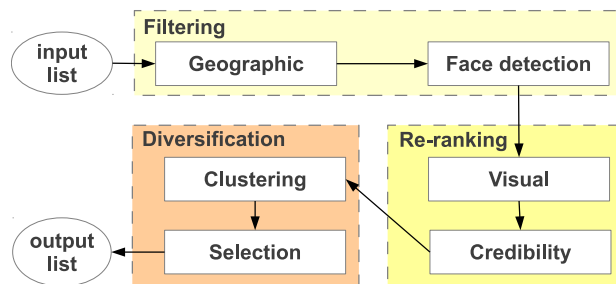
Figure 1: Proposed approach overview.

tagged images were assessed. According to the results on the development set, a 10 km range limit from the reference point was a good choice.

Since images containing a person or crowds in the foreground are considered non-relevant, we used a face detection module of Face++[1] for filtering. For all images, we computed the features: a) number of faces; b) biggest face size; c) smallest face size; d) average face size; e) total face size. The size values were computed as a fraction of the image spatial domain.

Our first face-based filtering approach (NumFacesFilter) eliminates all images with a number of faces superior to a threshold. According to the experiments on the development set, we eliminated all images with more than one face. The second approach (FaceClassifierFilter) used a 1-NN classifier based on the described features and considering all development images as training instances. All images classified as non-relevant were eliminated.

2.2 Features

For the textual and multimodal approaches, we evaluated the TF-IDF, BM25, and Cosine measures. All of them were computed using the provided TF, DF, and TF-IDF. In the development set, the best results were achieved using the Cosine measure. To enable the combination with other distance measures, the Cosine similarity values were converted by subtracting it from 1.0. For visual approaches, besides the provided features, we also extracted two global descriptors (BIC and LAS) [3] and two bag-of-visual-words (BoVW) descriptors, based on dense (6 pixels) or sparse (Harris-Laplace detector) SIFT, with 1000 visual words (randomly selected), soft assignment ($\sigma = 150$), and max pooling.

---

[1]http://www.faceplusplus.com - Last accessed on Sept 20, 2014.

2.3 Re-ranking

Since original lists may present redundant and non-relevant items, their positions may not be optimum for their relevance. Even after the filtering procedures, some non-relevant images may remain and therefore we proposed two re-ranking strategies: visual-based and user credibility-based.

The visual re-ranking used location's representative images obtained from Wikipedia as queries. The original lists were re-ranked according to the similarity in relation to the representative sets. The visual distance from each image in a list to the corresponding representative set was computed as the minimum distance value between the image and each representative image. For multiple feature fusion, we used a smoothed version of the Borda Count algorithm. In our version the vote (relevance score) for the nth image in the rank was computed as $\frac{1}{\sqrt[4]{n+1}}$.

As a different re-ranking strategy, we also exploited the user-credibility descriptors provided with the data. Hence, we combined a relevance-based score ($relScore$) with another score based on credibility ($credScore$). The $relScore$ of each image was computed according to its position in the list as described for the visual re-ranking. The $credScore$ was computed as the product of three credibility features: $visualScore$, $faceProportion$, and $tagSpecificity$. The final re-ranking score was computed as $relScore \times credScore$;

2.4 Diversification Method

After the filtering and re-ranking procedures, the next step consists of the actual summarization and diversification. We evaluated two diversification methods: MMR [1] and a clustering technique based-on kMedoids. Given the superiority of kMedoids over MMR on the development set, we used the kMedoids clustering for the test set runs.

The kMedoids clustering technique is divided into two main steps: medoid definition and clusters construction. Since we were supposed to return 50 representative images, the algorithm was set to create 50 clusters. The initial centroids were defined in a offset fashion. The offset value was computed by dividing the list size by 50. The centroids were then defined as the images in the positions $i \times$offset, with $0 \leq i < 50$. Hence, the initial medoids were picked throughout the list from the top to the bottom. After the clusters are constructed, the process iterates untill there is no further transition between the clusters. At each iteration, the new medoids were defined as the best connected images (average distance to all images in the cluster). The distance between two images is computed as the average of their distances computed for each feature. Finally the images in each cluster are ranked according to their positions in the original non-clustered list. The final output list is composed of the most relevant item from each cluster.

3. RUN SETUP

We submitted five runs and their descriptions are presented in Table 1. The features used in each run and each step were selected according to the best results on the development set.

4. RESULTS AND DISCUSSION

Table 4 presents the official evaluation measures for the five runs. We can see that best results (for all measures) were achieved when the proposed full pipeline was applied

Table 1: Runs Configurations

Run	Filtering	Re-ranking	Diversification
1	GeoFilter and NumFacesFilter	-	kMedoids ($BoVW_{max}^{sparse}$ + HOG)
2	GeoFilter and NumFacesFilter	-	kMedoids (Cosine)
3	GeoFilter and NumFacesFilter	Visual re-ranking (CM3x3 + HOG + BIC)	kMedoids ($BoVW_{max}^{sparse}$ + HOG + Cosine)
4	GeoFilter and NumFacesFilter	Visual re-ranking (CM3x3 + HOG + BIC) and Credibility re-ranking	kMedoids (CN3x3)
5	GeoFilter and FaceClassifierFilter	Visual re-ranking (CM3x3 + HOG + BIC) and Credibility re-ranking	kMedoids (CN3x3)

Table 2: Runs Effectiveness - Official Measures

Run	P@20	CR@20	F1@20
1	0.7130	0.4030	0.5077
2	0.6976	0.4139	0.5133
3	0.7016	0.4177	0.5168
4	**0.7598**	**0.4288**	**0.5423**
5	0.7407	0.4076	0.5206

(Runs 4 and 5). Run 2 (purely textual) slightly outperformed Run 1 (purely visual) in terms of diversity. The multimodal combination (Run 3) slightly outperformed Runs 1 and 2 on CR@20 and F1@20. However when the credibility re-ranking was applied (Run 4) the best results were achieved by the visual approach with reasonable improvement on all effectiveness measures. Notice that when the face-based filtering used the classifier (Run 5), the results were lower than using the face number threshold (Run 4) but still superior to Runs 1 to 3 on F1@20.

5. CONCLUSIONS

We proposed a multimodal approach with the use of filtering and re-ranking approaches in conjunction with a clustering technique for diversification. Our best results were achieved with image re-ranking by combining their relevance score and user credibility information. As future work we would like to evaluate the usage of additional information on the re-ranking and diversification steps and more elaborated fusion approaches.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] J. Carbonell and J. Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In SIGIR, pages 335–336, 1998.

[2] B. Ionescu, A. Popescu, M. Lupu, A. L. Gînscă, and H. Müller. Retrieving diverse social images at mediaeval 2014: Challenge, dataset and evaluation. In *MediaEval 2014 Workshop*, Barcelona, 2014.

[3] O. A. B. Penatti, E.Valle, and R. da S. Torres. Comparative study of global color and texture descriptors for web image retrieval. *J. Vis. Commun. Image Repr.*, 23(2):359 − 380, 2012.