Recod @ MediaEval 2015: Diverse Social Images Retrieval

Rodrigo T. Calumby[1,2], Iago B. A. do C. Araujo[2], Vinícius P. Santana[2],
Javier A. V. Munoz[1], Otávio A. B. Penatti[3], Lin T. Li[1], Jurandy Almeida[1,5],
Giovani Chiachia[1], Marcos A. Gonçalves[4], and Ricardo da S. Torres[1]

[1]RECOD Lab, University of Campinas, Campinas, SP – Brazil, [2]University of Feira de Santana, Feira de Santana, BA – Brazil
[3]SAMSUNG Research Institute Brazil, Campinas, SP – Brazil, [4]Federal University of Minas Gerais, Belo Horizonte, MG – Brazil
[5]GIBIS Lab, Federal University of São Paulo, São José dos Campos, SP – Brazil

[rtcalumby,ibacaraujo,vpsantana]@ecomp.uefs.br, jalvarm.acm@gmail.com, o.penatti@samsung.com,
jurandy.almeida@unifesp.br, mgoncalv@dcc.ufmg.br, [lintzyli,chiachia,rtorres]@ic.unicamp.br

ABSTRACT

This paper presents the RECOD team experience in the *Retrieving Diverse Social Images Task* at MediaEval 2015. The teams were required to develop a diversification approach for social photo retrieval. Our proposal is based on irrelevant image filtering, reranking, rank aggregation, and diversity promotion. We proposed a multimodal approach and exploited image metadata and user credibility information.

1. INTRODUCTION

The relevance-diversity trade-off is an important problem associated with several search scenarios. Promoting diversity in retrieval results has been shown to positively impact the user search experience specially for ambiguous, under-specified, and visual summarization queries [1]. The *Retrieving Diverse Social Images Task* [6] integrates it into a tourism related representative image retrieval challenge. This paper describes the RECOD group contributions via diversity promotion boosted by multimodal rank fusion.

2. PROPOSED APPROACH

As a relevance enhancement step our approach includes irrelevant image filtering, multimodal image reranking, and rank aggregation. Image filtering was conducted according to face detection data and geographic location of the images. We also evaluated several different visual features and text similarity measures. For reranking the original retrieval list, we exploited textual, visual, geographic, and credibility information. As an additional step, the reranked lists were also aggregated with a Genetic Programming (GP) [7] approach. Our proposal follows the general work-flow presented in Figure 1. Each of these steps is described next.

2.1 Filtering

For irrelevant images removal we used three filtering strategies: geographic-based, face-based, and blur-based. The GeoFilter eliminated all images located farther than a given range in relation to the reference lat/long. Following the result on the development set, the filtering radius was set to 10 km. In turn, the face-based procedure was introduced to filter out images containing people as the main subject. We
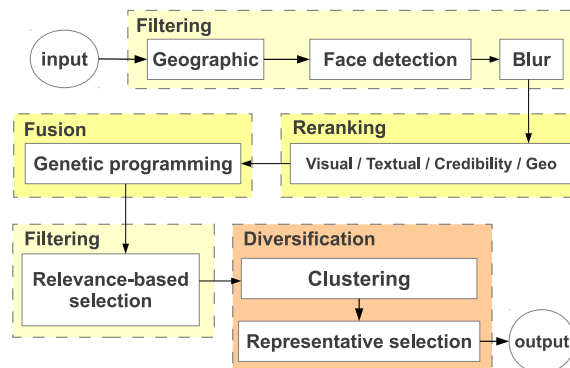
Figure 1: Overview of the proposed approach.

used a face detection module of Face++[1] and all images with more than one face detected were removed. Finally, out-of-focus images were eliminated using the method from [14] with blur threshold set to 0.8.

2.2 Visual Features and Text Similarity

For visual similarity, besides the provided features, we also extracted: (i) two general purpose global descriptors (BIC [13] and GIST [9]); (ii) a bag of visual words (BoVW) descriptor, based on sparse (Harris-Laplace detector) SIFT, with 512 visual words (randomly selected), soft assignment ($\sigma = 150$), and max pooling or using Word Spatial Arrangement (WSA) [11] for encoding the spatial arrangement of visual words; and (iii) fifteen features available in the Lire package [8].[2]

For text-only and multimodal runs, we used the Cosine, BM25, Dice, Jaccard, and TF-IDF measures which were computed using the provided TF, DF, and TF-IDF vectors.

2.3 Reranking and Aggregation

For improving the original list ranking, we explored visual, textual, credibility, and geographic ranking. For text-only and multimodal reranking, the text-based scores were computed as the similarity between the text vectors associated

---

[1]http://www.faceplusplus.com (as of June/2015).
[2]CEDD, FCTH, OpponentHistogram, JointHistogram, AutoColorCorrelogram, ColorLayout, EdgeHistogram, Gabor, JCD, JpegCoefficientHistogram, ScalableColor, SimpleColorHistogram, Tamura, LuminanceLayout, and PHOG. Available at: http://www.lire-project.net/

Table 1: Runs Configurations (*only for one-topic queries)

Run	Filtering	Reranking	Diversity
1	Geo*, face, blur	Visual*	BIC
2	Geo*, face, blur	Textual	Cosine + Jaccard
3	Geo*, face, blur	Visual*, textual	Jaccard
4	Geo*, face, blur	Credibility	Users
5	Geo*, face*, blur	Visual*, textual, credibility, geo*	Jaccard

Table 2: DevSet and TestSet Results

	DevSet			TestSet		
Run	P@20	CR@20	F1@20	P@20	CR@20	F1@20
1	0.7487	0.4336	0.5409	0.7129	0.4111	0.5063
2	0.8013	0.4514	0.5694	0.6996	0.4248	0.5101
3	0.7837	0.4436	0.5592	0.7058	0.3881	0.4883
4	0.7644	0.4446	0.5532	0.7198	0.4309	**0.5219**
5	0.8190	0.4637	**0.5853**	0.7324	0.4123	0.5084

Table 3: TestSet Results: One-topic and Multi-topic

	One-topic			Multi-topic		
Run	P@20	CR@20	F1@20	P@20	CR@20	F1@20
1	0.6906	0.4000	0.4991	0.7350	0.4221	**0.5133**
2	0.7130	0.4316	0.5205	0.6864	0.4181	0.4998
3	0.6942	0.3982	0.4970	0.7171	0.3782	0.4798
4	0.7630	0.4301	**0.5390**	0.6771	0.4318	0.5051
5	0.7290	0.4286	0.5228	0.7357	0.3963	0.4942

with the images and the localities' text vectors. The visual method reranked the original list according to the similarity in relation to the location's representative Wikipedia images. The visual distance from each image to the representative set was computed as the minimum distance to each representative image. All credibility scores were individually used for reranking. Additionally, lat/long data were used to rank images according to the Haversine distance to the reference point.

For feature fusion, the reranked lists were combined using the GP approach from [15] which uses several rank aggregation methods. This method was trained using the development data and combined order-based (MRA [4], RRF [3], and BordaCount [16]) and score-based (CombMIN, CombMAX, CombSUM, ComMED, CombANZ [12], and RLSim [10]) rank fusion methods.

As a relevance-based filtering, from the final aggregated list, up to 150-top ranked images were selected as the input list for the summarization procedure.

2.4 Diversification Method

After filtering, reranking, and aggregation steps, the improved relevance-based lists were submitted to explicit diversification. We evaluated four methods: clustering-based (k-Medoids and agglomerative) and reranking-based (MMR [2], MSD [5]). In all cases the k-Medoids method achieved significantly superior results on the development set and was used in the submitted runs.

In the clustering step, the initial medoids were selected in an offset fashion for equally sampling from the top to the bottom of the ranked list. The medoids updating procedure just selected the best connected image of the cluster, using the average distance to all other images in the cluster. The process iterates until no intercluster image transition occurs or up to 50 iterations.

For runs 1, 2, 3, and 5, the clusters were ranked according to their sizes in descending order and intra-cluster sorting was applied using average connectivity. For the credibility-based submission (run 4), the images were clustered according to their owner (user) and the clusters were ranked according to the users' credibility computed as a linear combination of tagSpecificity, uploadFrequency, meanTagRank, faceProportion, meanImageTagClarity, photoCount, visualScore, and locationSimilarity [6].

The representative images were selected in a round robin fashion from the final clusters. The number of clusters was defined as 30 for runs 1 and 3, and 40 for runs 2 and 5.

3. RUNS SETUP

We submitted five runs (Table 1). The features used for diversification were selected according to the best results on the development set.

In all runs, the geographic filtering and reranking were only applied for one-topic queries since multi-topic queries do not have reference geo-location. Additionally, in run 5, the face-based filters were also only applied to one-topic queries since multi-topic queries have a different relevance constraint in relation to people in the foreground. Finally, in runs 1, 3, and 5 no visual reranking was applied for multi-topic queries.

4. RESULTS AND DISCUSSION

Table 2 presents the effectiveness results for the five runs for the development and test set. The best results (F1@20) on the development set were achieved on run 5, followed by runs 2 and 3, in which textual information was used. However, these were the runs with the greatest effectiveness difference when comparing development and test queries, specially considering the multi-topic queries.

Table 3 presents the effectiveness results for one-topic and multi-topic test queries. As we can observe, even with no visual reranking, the visual-only run allowed slightly superior results for multi-topic queries considering all the run types and also comparing to one-topic queries. All other run types achieved superior effectiveness on one-topic queries, specially when the credibility information was used (runs 4 and 5). Our results suggests that visual features are important when considering multi-topic queries while the textual information seems more suitable for one-topic queries.

5. CONCLUSIONS

For relevance and diversity maximization we proposed filtering strategies and the combination of multiples features with a rank fusion method. These improved ranked lists were used as input for a clustering-based summarization method. Our experiments suggest that different summarization alternatives may result in different effectiveness for one-topic and multi-topic queries.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] R. T. Calumby, R. da S. Torres, and M. A. Gonçalves. Diversity-driven learning for multimodal image retrieval with relevance feedback. In *Proceedings of the 21st IEEE International Conference on Image Processing*, pages 2197–2201, 2014.

[2] J. Carbonell and J. Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Anual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 335–336, 1998.

[3] G. V. Cormack, C. L. A. Clarke, and S. Buettcher. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 758–759, 2009.

[4] R. Fagin, R. Kumar, and D. Sivakumar. Efficient similarity search and classification via rank aggregation. In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, pages 301–312, 2003.

[5] S. Gollapudi and A. Sharma. An axiomatic approach for result diversification. In *Proceedings of the 18th International Conference on World Wide Web*, pages 381–390, 2009.

[6] B. Ionescu, A. Gînscă, B. Boteanu, A. Popescu, M. Lupu, and H. Müller. Retrieving diverse social images at mediaeval 2015: Challenge, dataset and evaluation. In *Working Notes Proceedings of the MediaEval 2015 Workshop*, Wurzen, Setember 14-15 2015. CEUR-WS.org.

[7] J. R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 1992.

[8] M. Lux and S. A. Chatzichristofis. LIRE: lucene image retrieval: an extensible java CBIR library. In *Proceedings of the 16th ACM International Conference on Multimedia*, pages 1085–1088, 2008.

[9] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. 42(3):145–175, 2001.

[10] D. C. G. Pedronette and R. da S. Torres. Image re-ranking and rank aggregation based on similarity of ranked lists. In *Proceedings of the 14th International Conference on Computer Analysis of Images and Patterns - Volume Part I*, pages 369–376, 2011.

[11] O. A. B. Penatti, F. B. Silva, E. Valle, V. Gouet-Brunet, and R. da S. Torres. Visual word spatial arrangement for image retrieval and classification. *Pattern Recognition*, 47(2):705–720, 2014.

[12] J. A. Shaw, E. A. Fox, J. A. Shaw, and E. A. Fox. Combination of multiple searches. In *The Second Text REtrieval Conference (TREC-2)*, pages 243–252, 1994.

[13] R. Stehling, M. Nascimento, and A. Falcão. A compact and efficient image retrieval approach based on border/interior pixel classification. In *Proceedings of the 11th International Conference on Information and Knowledge Management*, pages 102–109, 2002.

[14] H. Tong, M. Li, H. Zhang, and C. Zhang. Blur detection for digital images using wavelet transform. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 17–20 Vol.1, 2004.

[15] J. A. Vargas, R. da S. Torres, and M. A. Gonçalves. A soft computing approach for learning to aggregate rankings. In *Proceedings of the 24th ACM International Conference on Conference on Information and Knowledge Management*, 2015. (in press).

[16] H. P. Young. An axiomatization of borda's rule. *Journal of Economic Theory*, 9(1):43–52, 1974.