Recod @ MediaEval 2016: Diverse Social Images Retrieval

Cristiano D. Ferreira[1], Rodrigo T. Calumby[2], Iago B. A. do C. Araujo[2]
Ícaro C. Dourado[1], Javier A. V. Munoz[1], Otávio A. B. Penatti[1,3], Lin T. Li[1]
Jurandy Almeida[1,4] 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]GIBIS Lab, Federal University of São Paulo, São José dos Campos, SP – Brazil
[rtcalumby,ibacaraujo]@ecomp.uefs.br, [crferreira,icarocd,jalvarm.acm]@gmail.com, o.penatti@samsung.com,
jurandy.almeida@unifesp.br, [lintzyli,rtorres]@ic.unicamp.br

ABSTRACT

This paper presents the RECOD team experience in the *Retrieving Diverse Social Images Task* at MediaEval 2016. The teams were required to develop a diversification approach for social photo retrieval. Our proposal is based on re-ranking, rank aggregation, and diversity promotion, allowing employment of textual and visual information apart or fused.

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, underspecified, and visual summarization queries [2].

The *Retrieving Diverse Social Images Task 2016* [7] task addresses the problem of image search result diversification in the context of social media. This paper describes the RECOD group contributions via diversity promotion boosted by rank fusion.

2. PROPOSED APPROACH

Our proposal follows the general workflow presented in Figure 1. The first step, *Re-ranking*, ranks the original list provided by Flickr according to a text-based descriptor. The *Fusion* step employs Genetic Programming (GP) [8] to aggregate lists re-ranked by several text-based descriptors. Finally, the *Diversification* step exploits visual and textual-based descriptors to promote diversification at the resulting ranking.

The next sections provide a more detailed description of our approach.

2.1 Visual Features and Text Similarity

For visual similarity, besides the provided features, we also extracted: (i) two general-purpose global descriptors (BIC [15] and GIST [11]); (ii) a bag-of-visual-words (BoVW) descriptor, based on sparse (Harris-Laplace detector) SIFT, with 1000 visual words (randomly selected), soft assignment ($\sigma = 150$), and max pooling with spatial pyramids or Word Spatial Arrangement (WSA) [13] for encoding the spatial arrangement of visual words; and (iii) fifteen features available

MediaEval 2016 Workshop, October 20-21, 2016, Hilversum, Netherlands.

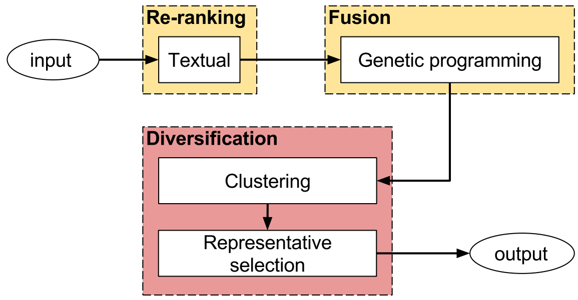

Figure 1: Overview of the proposed approach.

in the Lire package [10].[1]

For text-only and multimodal runs, we used the cosine [1], BM25 [1], Dice [9], Jaccard [9], and TF-IDF measures which were computed using the provided TF, DF, and TF-IDF vectors.

2.2 Re-ranking and Aggregation

For improving the original list ranking, several textual measures (Section 2.1) were employed for re-ranking. The text-based scores were computed as the similarity between the text vectors associated with the query topic and the image associated text vectors. For visual only run, the re-ranking step was skipped.

For feature fusion, re-ranked lists were combined using the GP approach from [16], which uses several rank aggregation methods. This method took as input the Flickr query result re-ranked considering each textual similarity measures. Then, it was trained using the development data and combined by order-based (MRA [5], RRF [4], and BordaCount [17]) and score-based (CombMIN, CombMAX, CombSUM, ComMED, CombANZ [14], and RLSim [12]) rank fusion methods.

2.3 Diversification Method

After re-ranking and aggregation steps, the improved relevance-based lists were submitted to explicit diversifica-

[1]CEDD, FCTH, OpponentHistogram, JointHistogram, AutoColorCorrelogram, ColorLayout, EdgeHistogram, Gabor, JCD, JpegCoefficientHistogram, ScalableColor, SimpleColorHistogram, Tamura, LuminanceLayout, and PHOG. Available at: `http://www.lire-project.net/` (As of Sep. 2016).

tion. Visual and textual descriptors were employed (Section 2.1). We evaluated five methods: clustering-based (k-Medoids, agglomerative and Birch [18]) and re-ranking-based (MMR [3] and MSD [6]).

Agglomerative and Birch methods achieved significantly superior results on the development set, thus they were used in the submitted runs.

In the clustering step, for agglomerative method, centroid and average link linkage methods were employed using ColorLayout for distance computing. Forty clusters were considered in our approach. For Birch method, a maximum of 51 entries per node was admitted, with a distance threshold of 0.3, and also considering cluster refining. The representative images were selected in a round robin fashion from the final clusters.

2.4 Workflow Discussions

It is important to observe that re-ranking and GP fusion are optional steps at the workflow presented in Figure 1. Depending on run requirements or on the desired experiment goals, one or both can be skipped. For example, for run 1, which is visual only, neither of those steps were employed.

Furthermore, the diversification step can employ textual or visual information alone or together. It allows adherence to the single modality requirement of runs 1 and 2. Section 3 will present details of each run configuration.

3. RUNS SETUP

We submitted five runs.

Run 1 – *(required) visual information only.* No re-ranking and no GP-fusion were employed. Diversification provided by Agglomerative method, using average link method with ColorLayout as visual feature and grouping images into 40 clusters.

Run 2 – *(required) text information only.* Re-ranking considering cosine similarity was employed. Diversification provided by Birch method, using 51 as maximum entries per node with distance threshold of 0.3.

Run 3 – *(required) text-visual fused.* Re-ranking considering cosine similarity was employed. Diversification provided by Birch method, using 51 as maximum entries per node with distance threshold of 0.19;

Run 4 – *(optional) general run.* GP-fusion rank aggregation employed (Figure 2 - a). Diversification provided by Agglomerative method, using average link method with ColorLayout as visual feature and grouping images into 40 clusters;

Run 5 – *(optional) general run.* GP-fusion rank aggregation employed (Figure 2 - b). Diversification provided by Agglomerative method, using centroid linkage method with ColorLayout as visual feature and grouping images into 40 clusters.

The diversification methods, parameters, features, and textual similarities used were selected according to the best results on the development set.

4. RESULTS AND DISCUSSION

Table 1 presents the results for the five runs for the development and test sets. The best results (F1@20) on the development set were achieved on run 2, followed by runs 1 and 3, in which textual information was used. Runs 4 and 5, which employed GP-fusion rank aggregation, have shown

```
CombMAX(
 BordaCount(
  CombMAX(
   CombMNZ(COSINE, BM25),
   COSINE_ME
  ),
  CombMAX(
   CombMNZ(DICE, DICE),
   COSINE
  )),
 CombMIN(
  RRF(
   CombMNZ(DICE, TFIDF),
   COSINE_ME
  ),
  CombMAX(
   MRA(TFIDF, BM25Orig),
   COSINE
  )))
```

(a)

```
MRA(
 CombMNZ(
  CombMAX(
   CombSUM(BM25, COSINE),
   RRF(JACCARD, COSINE)),
  BM25),
 MRA(TFIDF, COSINE_ME))
```

(b)

Figure 2: GP individuals for rank aggregation.

Table 1: DevSet and TestSet Results

	DevSet			TestSet		
Run	**P@20**	**CR@20**	**F1@20**	**P@20**	**CR@20**	**F1@20**
1	**0.6821**	0.4641	0.5359	**0.5180**	0.4001	0.4258
2	0.6814	**0.4643**	**0.5403**	0.5000	0.3709	0.4045
3	0.6714	0.4519	0.5268	0.4969	0.3881	0.4107
4	0.6614	0.4578	0.5248	0.5156	**0.4173**	0.4339
5	0.6550	0.4525	0.5196	0.5156	0.4065	**0.4379**

the worst results on this set. However, on the test set, they presented the best results.

As we can observe, in most cases, the re-ranking over the Flickr initial ranking improved the overall results, even when employing only textual features for this task.

Considering test set results of runs 4 and 5 over 1 and 2, we can also notice that the retrieval process illustrated in Figure 1 can benefit from fusing visual and textual information. We believe that textual and visual information have a complementary nature for an effective image retrieval process. Textual information introduces the notion of context around retrieval, but ignores the image itself by not inspecting its content. On the other hand, content-based image retrieval lacks context. By initial retrieval and re-ranking based on textual information and introducing visual features at the diversification step this complementary nature is explored and provided the best results.

5. CONCLUSIONS

For relevance and diversity maximization, we proposed re-ranking strategies and the combination of multiple features with a rank fusion method. These improved ranked lists were used as input for a clustering-based summarization method. Our experiments suggest that aggregation of multiple re-ranked lists and fusion of visual and textual information can improve retrieval effectiveness. It is interesting to observe the recurrent selection of the cosine measure on both GP aggregation individuals.

Acknowledgments

We thank the support of CAPES, CNPq, and FAPESP.

6. REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval - the concepts and technology behind search, Second edition*. Pearson Education Ltd., Harlow, England, 2011.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] B. Ionescu, A. L. Gînscă, M. Zaharieva, B. Boteanu, M. Lupu, and H. Müller. Retrieving diverse social images at mediaeval 2016: Challenge, dataset and evaluation. In *Proceedings of the MediaEval 2016 Workshop*, Hilversum, Netherlands, Oct. 20-21 2016.

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

[9] J. Lewis, S. Ossowski, J. Hicks, M. Errami, and H. R. Garner. Text similarity: an alternative way to search medline. *Bioinformatics*, 22(18):2298–2304, 2006.

[10] 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.

[11] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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 on Conference on Information and Knowledge Management*, CIKM '15, pages 83–92, New York, NY, USA, 2015. ACM.

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

[18] T. Zhang, R. Ramakrishnan, and M. Livny. Birch: An efficient data clustering method for very large databases. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, SIGMOD '96, pages 103–114, New York, NY, USA, 1996. ACM.