

CFM@MediaEval 2017 Retrieving Diverse Social Images Task via Re-ranking and Hierarchical Clustering

Liang Peng, Yi Bin, Xiyao Fu

Jie Zhou, Yang Yang, Heng Tao Shen

Center for Future Media and School of Computer Science and Engineering

University of Electronic Science and Technology of China

pliang951125@outlook.com, yi.bin@hotmail.com, fu.xiyao.gm@gmail.com

jiezhou0714@gmail.com, dlyyang@gmail.com, shenhengtao@hotmail.com

ABSTRACT

This paper presents an approach based on re-ranking and hierarchical clustering (HC) for *MediaEval 2017 Retrieving Diverse Social Images Task*. The experimental results on the development and test set demonstrate that the proposed approach can significantly improve relevance and visual diversity of the query results. Our approach achieves a good tradeoff between relevance and diversity and a result in F1@20 of 0.6533 for the employed test data.

1 INTRODUCTION

Modern image retrieval systems should be able to give both relevant and visually diverse results. In other words, the presented images should not only be relevant to the query, but also cover different visual facets. The *MediaEval 2017 Retrieving Diverse Social Images Task* [8] fosters the development of such image retrieval systems, which aim at the improvement of both the relevance and the diversity of image search results. More specifically, given images retrieved from Flickr using text queries, the goal of the task is to refine the results by ranking images according to their relevance to the query and offering visually diversity. More details about the task and data description can be found in [8].

Previous participants, such as [3, 7] employed a re-ranking method to improve the relevance and showed a promising performance. Clustering algorithms were widely used in this task to cluster similar images, such as k-Medoids [4], hierarchical clustering [1, 2], random forest [6], etc.

In this work, we propose to re-rank images using the available social metadata in order to refine the original relevance rank list of images provided by Flickr. Then, a hierarchical clustering (HC) algorithm is employed to diversify the top k images. Unlike Tolari [7], who merges different features based on similarities, we merge features using weighted distances. In order to balance relevance and diversity well, we apply a diversification strategy similar to [1].

2 APPROACH

As shown in Figure 1, our approach is composed of three components: *Re-ranking*, *Clustering* and *Diversification*. Specifically, to improve the semantic relevance between images and their corresponding queries, we first re-rank the images retrieved for each

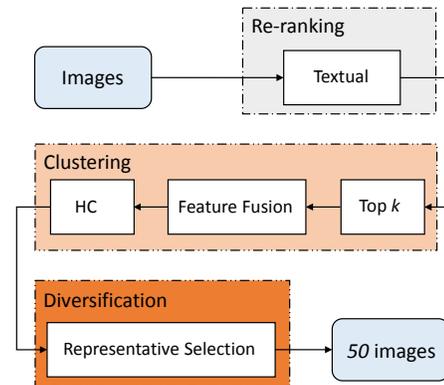


Figure 1: Overview of our approach.

topic using the available textual metadata. Then we apply a hierarchical clustering algorithm to diversify the refined image rank list and balance relevance and diversity of the final results using a diversification approach.

2.1 Re-ranking

The given dataset consists of images that are retrieved from Flickr using text-based, multi-topic queries. For each topic, images are arranged using the initial rank list by the Flickr retrieval system. The rank list, however, commonly mismatches the probability of relevance to its query topic. To improve the relevance of the initial rank list, we propose to re-rank the images of each topic by exploiting their social metadata (e.g., *titles*, *geo-tags*, and *usernames*). Concretely, we compute the Cosine Similarity according to the TF-IDF weights of the combined social metadata (i.e., concatenating corresponding feature vectors). Note that we do not utilize the *descriptions* in metadata, because we find that there is much redundant or irrelevant information in the *descriptions*, which may decrease the relevance of the underlying images and the final performance notably. Since the queries are represented by text, our re-ranking could narrow the semantic gap between the queries and the relevance of the corresponding image results.

2.2 Clustering

After re-ranking, we obtain a more relevant rank list of the initial N images for each topic ($N = 300$ in the task). For the purpose of reducing irrelevant images, we only select the top k images from the refined rank list for the hierarchical clustering (See Section 4

Table 1: Run performance on Devset

	run1	run2	run3	run4	run5	<i>Flickr</i>
P@20	0.55	0.6473	0.6595	0.6509	0.6605	0.5864
CR@20	0.3959	0.453	0.4768	0.4838	0.4888	0.3646
F1@20	0.4409	0.5152	0.5323	0.534	0.5402	0.4277

for more details). In other words, the images with low ranks, which are assumed to be the most irrelevant to the topic, are discarded. The experimental results indicate that the employed performance scores and F1@20 benefit from this selective operation.

At the clustering step, one or more visual features, textual features and credibility descriptors are utilized. In order to take advantage of several features, we merge them by summing the weighted distance:

$$dist_{f_1, f_2, \dots, f_n}(I_1, I_2) = \sum_{i=1}^n w_i \cdot dist_{f_i}(I_1, I_2), \quad (1)$$

where f_i denotes the i -th feature, $dist_{f_i}(I_1, I_2)$ is the distance between image 1 and image 2 based on i -th feature, and w_i is a manually set weight for the distance of the i -th feature.

We employ HC with single, complete, ward and average linkage methods [5] respectively based on the Euclidean distance, and cluster 50 clusters for each topic.

2.3 Diversification

HC distributes the selected top k images to several sets, each of which potentially covers one scene or situation with related semantic meaning. Intuitively, to improve the diversity of a retrieval system, the results should cover as many clusters as possible. In the similar way, the higher ranks indicate higher relevance. To balance diversity and relevance, we directly select the images with the highest relevance ranks from each cluster and sort them according to their relevance as final result (see re-ranking in Section 2.1).

3 RUN SETUP

We submit 5 runs with different experimental setting.

- Run1: *visual-only*. No re-ranking. HC uses a complete linkage method and the provided visual feature *auto color correlogram* (*acc*) [8].
- Run2: *text-only*. Text-based re-ranking. HC uses a complete linkage method and the provided textual feature [8].
- Run3: *text-visual*. Text-based re-ranking. HC uses a ward linkage method and the provided visual feature *scalable color* (*sc*) and textual feature [8].
- Run4: *text-cred*. Text-based re-ranking. HC uses a single linkage method and the provided credibility descriptors [8].
- Run5: *text-cred*. Text-based re-ranking. HC uses a single linkage method and the provided credibility descriptors and textual feature [8].

4 EXPERIMENT AND RESULTS

Table 1 and Table 2 present the results for the five runs evaluated on Devset and Testset respectively. Concretely, Devset is the development set and Testset is the official test set. Results are assessed

Table 2: Run performance on Testset

	run1	run2	run3	run4	run5
P@20	0.628	0.6833	0.6881	0.6863	0.6881
CR@20	0.6009	0.6274	0.6561	0.6636	0.6671
F1@20	0.5871	0.6334	0.6488	0.6505	0.6533

with Precision at X images (P@X), Cluster Recall at X (CR@X), and F1-measure at X (F1@X). F1@20 is the official ranking metric. Best results are depicted in bold. *Flickr* is the baseline result using the top 50 images based on the initial Flickr ranking.

In run1, we do not use the top k strategy before clustering because of no re-ranking. Our preliminary experiments show that *acc* can achieve better performance than other provided visual features (e.g., *CNN*, *sc*) in run1. The performance of run1 in Table 1 shows that clustering with visual feature *acc* can improve the diversity to a certain degree. In run2, we only employ textual features (*title*, *tags*, *username*) for both re-ranking and clustering. Before clustering, we select the top k (100, 150, 200, 250, 300) images. Preliminary experiments indicate that $k=150$ achieves the best performance in terms of F1@20. The results achieved on the development and test data show that re-ranking and hierarchical clustering using textual features can improve the relevance and the diversity in comparison to run1 and *Flickr* to a large degree.

The core difference between run2, run3, run4 and run5 is that they employ different features for clustering. Run3 is a combined visual-textual run and employs a ward linkage method for clustering. The reason we do not use the same visual feature as run1 is that better performance will be achieved using *sc* than *acc* or any other provided visual features in run3. The weights of textual and visual feature are 1 and 0.02 respectively. In contrast to the result of run2 on Devset and Testset, the value of every metric is increased, CR@20 especially. The results on Devset and Testset indicate that the combination of visual and textual features outperforms the single feature. Note that run4 which employs only credibility feature in clustering achieves a better CR@20 and F1@20 than run2 and run3 on both Devset and Testset. In our experiments, the credibility feature is the best feature for clustering. Run5 combines credibility and textual features in clustering and achieves the best performance in P@20, CR@20 and F1@20 on both Devset (0.6605, 0.4888, 0.5402) and Testset (0.6881, 0.6671, 0.6533). We do not use any visual feature, because our experimental results on the development set show that none of the provided visual features is helpful to attain a higher F1@20 in run5. The weights of textual and credibility features are 0.2 and 1 respectively.

5 CONCLUSION

We presented an approach which employs re-ranking and hierarchical clustering for retrieving diverse social images. Experimental results demonstrated the efficiency of our approach to improve the relevance and diversity. Credibility descriptors provided a mark of the quality of the user’s tag-image content relationships and were very helpful to cluster similar images to improve the diversity. Merging textual and credibility features could take advantage of different features and achieved satisfactory performance.

REFERENCES

- [1] B. Boteanu, M.G. Constantin, and B. Ionescu. Lapi@2016 retrieving diverse social images task: A pseudo-relevance feedback diversification perspective. In *MediaEval*, 2016.
- [2] X. Chen, L. Haokun, D. Zhi-Hong, and Y. Yunlun. Retrieving relevant and diverse image from social media images. In *MediaEval*, 2015.
- [3] C. D Ferreira, R. T. Calumby, I. B. do C Araujo, Ícaro C Dourado, J. AV Munoz, O. A. B. Penatti, L. T. Li, J. Almeida, and R. da Silva Torres. Recod@ mediaeval 2016: Diverse social images retrieval. In *MediaEval*, 2016.
- [4] A. L. Gînsca, A. Popescu, and N. Rekabsaz. Cea list's participation at the mediaeval 2014 retrieving diverse social images task. In *MediaEval*, 2014.
- [5] D. Müllner. Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*, 2011.
- [6] C. Spampinato and S. Palazzo. Perceive lab@unict at mediaeval 2014 diverse images: Random forests for diversity-based clustering. In *MediaEval*, 2014.
- [7] S. Tollari. Upmc at mediaeval 2016 retrieving diverse social images task. In *MediaEval 2016 Workshop*, 2016.
- [8] M. Zaharieva, B. Ionescu, A. L. Gînsca, R. L. T. Santos, and H. Müller. Retrieving diverse social images at mediaeval 2017: Challenge, dataset and evaluation. In *MediaEval*, 2017.