CEA LIST's Participation at the MediaEval 2014 Retrieving Diverse Social Images Task

Alexandru Lucian Ginsca[1,2], Adrian Popescu[1], Navid Rekabsaz[3]
[1]CEA, LIST, Vision & Content Engineering Laboratory, 91190 Gif-sur-Yvette, France
[2]TELECOM Bretagne, France
[3]Faculty of Informatics, Vienna University of Technology
{alexandru.ginsca, adrian.popescu}@cea.fr, rekabsaz@ifs.tuwien.ac.at

ABSTRACT

The Mediaeval 2014 Retrieving Diverse Social Image Task aims to tackle the challenge of improving result diversity while keeping a high precision in a social image retrieval task. We base our approach on the retrieval performance of recently introduced visual descriptors coupled with a mixt diversification method that explores the use of social cues together with a classic clustering setting. As a novelty, this year's task introduced user credibility features. We also describe how to use credibility in the diversification process and how to improve individual features by the means of a regression model.

1. INTRODUCTION

Social image retrieval presents an appropriate setting for the use of multimodal approaches to improve both results relevance and diversity. Recently, emerging works propose the use of social cues alongside visual and textual data.

Our efforts are channeled towards exploiting visual information and the use of credibility in the diversification process. We first describe a couple of pre-filtering techniques followed by an image retrieval method that boosts precision. Next, we describe how to predict a user's credibility score and we propose a user based image filtering approach. After we show how we improve diversity by clustering and cluster ranking, we finally describe the submitted runs and discuss the results we obtained on the testset.

2. AIMING FOR PRECISION

2.1 Initial pre-filtering

We use two filtering steps with the goal to eliminate noise form the image lists. Similar to [2], we eliminate geotagged images that have a distance from the POI higher than 1 km. The second filter is a restriction on the presence of faces in images. We use the standard OpenCV[1] algorithm to perform face detection and we eliminate images having a face coverage ratio higher than 0.4. The distance threshold and the one for the percentage of faces are determined on the devset. We keep the same pre-filtering steps for all the runs.

2.2 Image retrieval

Following the latest advances in computer vision, we use Caffe [3], a powerful CNN-based feature, to extract representations for the images in the collection, as well as the Wikipedia image examples. Following a standard content based image retrieval approach, we rank the images for each topic by the average cosine similarity between the retrieved image and all of the example images. On the devset, we obtain a P@20 of 0.966 when doing retrieval with the Caffe features. This represents a significant improvement over the Flickr ranking (P@20 = 0.831) and LBP3x3 (P@20 = 0.816), the descriptor provided by the organizers which gives the best performances in visual retrieval. One drawback of this method is the strong trade-off between precision and cluster recall. Although P@20 on the devset is high, we get a CR@20 of 0.293, leading to a F1@20 of 0.438. This problem is directly approached by first selecting images found in different clusters, as described in Section 4.

3. LISTENING TO SOCIAL CUES

3.1 Predicting user credibility

We exploit the credibility set to train a regression model that predicts a user's credibility score from the provided features. We perform model selection and parameter tunning by 5-fold cross-validation (cv) on the credibility set and we evaluate the performance of the predictions by Spearman's rank correlation coefficient with the ground truth credibility values. The highest cv correlation (0.47) is obtained using gradient boosting regression trees with a Huber loss and 100 estimators. By comparison, the highest correlation of an individual feature (*visual score*) is 0.36. The gain in regards to the Spearman score is also reflected on the competition metrics. When fixing the rest of the parameters and using the predicted credibility scores instead of the provided visual credibility feature, F1@20 increases from 0.61 to 0.632 on the devset.

3.2 User selection

For each topic, we first keep a subset of users that have contributions in the top n images found in the ranking produces by the image retrieval process described in Section 2.2. Then, as an extra filter, in our final ranking we retain only images coming from the selected user set. Given the good precision of image retrieval, we have a high confidence that images found in the top of the ranking are relevant. This gives us an ad-hoc topical expertise insight about the users responsible for those images. We tune n on the devset and

[1] http://opencv.org/

MediaEval 2014 Workshop, October 16-17, 2014, Barcelona, Spain

fix it at 20. For comparison, when not using a user based filter, the F1@20 score drops from 0.632 to 0.597. We also tried a similar approach by retaining contributions from top users ranked according to the credibility score but this did not improve the results. This result hints at the need for a topic specific credibility score.

4. IMPROVING DIVERSITY

Building on previous works, we combine a more traditional clustering approach for diversification with the use of social cues [5].

4.1 Clustering

We first perform k-Means clustering on the complete set of images. To ensure a stable cluster distribution, we initialize the centroids by uniformly selecting images from the ranking produced after image retrieval. For example, the i-th cluster will have as initial centroid the image found on the position $(i-1) * n/k$, where k is the desired number of clusters and n is the number of images in the ranking. After validation on the devset, k is set to 30.

4.2 Cluster ranking

We leverage the social component of this task by ordering the clusters based on the average credibility score of the users that contribute with images in the cluster. For the runs that do not permit the use of credibility, we rank the clusters according to the number of unique users represented in each cluster. In the case of a tie, we prefer the cluster that has the best ranked image after visual retrieval. Our final ranked list is obtained by selecting from each cluster at a time the image that is best placed in the visual retrieval ranking.

5. RESULTS AND DISCUSSION

We submitted five different runs at this year's Retrieving Diverse Social Images Task [1]. Our submissions are briefly described below:

- RUN1 uses the provided LBP3x3 visual descriptor for image retrieval and clustering. The clusters are then ranked based on the number of users represented in each cluster.
- RUN2 is a purely textual one. We concatenated the title, tags and description of the photos to calculate the text similarity. As text pre-processing phase, we decompounded the terms by applying a greedy approach using the dictionary which is created by all the words in the text. In the next step, in order to disambiguate the places, we expand the queries using the first sentence of Wikipedia. After testing several language models, using a semantic similarity approach based on Word2Vec [4] gave the best result. We trained a model on Wikipedia and then used the vector representation of words to calculate the text similarity of the query to each photo. In additional to the text similarity, we extracted three binary attributes: (1) if the photo had any views, (2) if the distance between a photo and the POI is greater than 8 kilometers, and (3) if the description length has more than 2000 characters. All features were then used in a Linear Regression model in order to re-rank the list. Finally, following [5], in order to diversify the ranking, we iterate over the initial re-ranked list and keep one image from each user at each iteration.
- RUN3 is a fusion between RUN1 and RUN2. Since the scores for visual and textual rankings are not in the same range, fusion is performed based on the ranks of the images in the two initial rankings. More specifically, we perform a linear weighting in which the individual ranks are given a weight of 0.5. Other weighting have been tested but the results remain quite stable in the range 0.3 - 0.7, a result which accounts for the robustness of the proposed fusion.
- RUN4 is similar to RUN1 with the single difference laying in the use of credibility for cluster ranking.
- RUN5 is obtained using the Caffe visual descriptor for image retrieval and clustering and predicted credibility scores for cluster ranking.

Table 1: Run performances with three official metrics

Run name	F1@20	P@20	CR@20
RUN1	0.5182	0.7313	0.4103
RUN2	0.5346	**0.8089**	0.4084
RUN3	0.5525	0.798	0.4335
RUN4	0.5243	0.7378	0.4157
RUN5	**0.571**	0.7931	**0.4563**

Our textual run (RUN2) is the single one in which we do not use clustering to improve diversity. This reflects across metrics, as it can be seen in Table 1. Although it performs well in terms of F1@20, this run is placed at oposite poles when looking at the other metrics. It has the highest P@20 and the lowest CR@20.

The usefulness of credibility can be best observed when comparing RUN1 and RUN4. They share the same configuration with the sole exception being the use of the predicted credibility scores for cluster ranking in RUN4. Although the difference is not as significant as on the devset, we can see a slight improvement of F1@20.

6. ACKNOWLEDGMENT

This research was supported by the MUCKE project, partly funded within the FP7 CHIST-ERA scheme.

7. REFERENCES

[1] B. Ionescu and al. Retrieving diverse social images at mediaeval 2014: Challenge, dataset and evaluation. In *MediaEval 2014 Workshop*, Barcelona, Spain, October 16-17 2014.

[2] N. Jain and al. Experiments in diversifying flickr result sets. In *MediaEval 2013 Workshop*, Barcelona, Spain, October 18-19 2013.

[3] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. *http://caffe.berkeleyvision.org*, 2013.

[4] T. Mikolov and al. Efficient estimation of word representations in vector space. *CoRR*, 2013.

[5] A. Popescu. Cea list's participation at the mediaeval 2013 retrieving diverse social images task. In *MediaEval 2013 Workshop*, Barcelona, Spain, October 18-19 2013.