

DCLab at MediaEval2015 Retrieving Diverse Social Images Task

Zsombor Paróczy
Budapest University of
Technology and Economics
paroczi@tmit.bme.hu

Kis-Király Máté
Budapest University of
Technology and Economics
kis.kiraly.mate@gmail.com

Dániel Mira
Budapest University of
Technology and Economics
miradaniellevante@gmail.com

ABSTRACT

In this paper we recommend a social image re-ranking method for the MediaEval 2015 Retrieving Diverse Social Images Task in order to increase the accuracy of a search result on Flickr based on relevance and diversity. Our approach is based on reranking the original result, using precalculated distance matrix to increase diversity. We have used color related visual features, text and credibility descriptors to define similarity between images.

1. INTRODUCTION

When a potential tourist makes a search on a place, he (or she) wants to have a diverse and relevant visual result as a summary of the different views of the location.

In the official challenge (Retrieving Diverse Social Images at MediaEval 2015: Challenge, Dataset and Evaluation) [2] a ranked list of location photos retrieved from Flickr is given, and the task is, to refine the results by providing a set of images that are both relevant and provide a diversified summary. The diversity means that images can illustrate different views of the location at different times of the day/year and under different weather conditions, creative views, etc. The goodness of the refinement process can be measured using the precision and diversity metric [?].

Our team participated in previous challenges [7, 6], each year we experimented with a different approach. In 2013 we used diversification of initial results using clustering, but our solution was focused on diversification only. In 2014 we tried to focus on relevance and diversity with the same importance, as a new idea.

In this paper and in this year's challenge we used a more sophisticated distance based clustering method, where we crafted the distance matrix ourselves rather than using some n dimension based clustering with an euclidean distance function.

2. RUNS

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MediaEval 2015 Workshop, Sept. 14-15, 2015, Wurzen, Germany
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

2.1 Run1: Visual based reranking

In the first subtask participants could use only visual based metrics or own metrics calculated using only the images.

Our main idea was using color based distances [1, 5] and filtering photos with faces on them [7, 6]. We experimented with HOG feature distances, but didn't achieve any additional improvement.

First we calculated a new value to the image: the *FACE* feature, which is the ratio of the faces on an image [7]. Then we used the *CN* metric to filter out black color based images, since mostly dark images tend to have less colors and those are mainly shifted into the gray region rather than having bright colors.

In the reordering step we started from the original ordering. We did our initial filtering by putting images to the end of the result list where $FACE = 0.0$ or $CN[0] > 0.8$. With this step we removed the images with people on it (e.g. family photos) and dark images (e.g. fireworks).

After the preprocessing steps we build the distance D matrix between each A and B images is calculated using the following equation:

$$D_{A,B} = \sum_{i=0}^{10} |CN_A[i] - CN_B[i]| + \sum_{i=0}^{10} |s_i * (CM_A[i] - CM_B[i])|$$
$$s_i = \begin{cases} 5, & \text{where } 0 \leq i < 3 \\ 1.5, & \text{where } 3 \leq i < 5 \\ 0.5, & \text{where } 5 \leq i < 9 \end{cases}$$

After the distance matrix was created we used spectral clustering [3, 4] to create clusters from the first 150 images, the target cluster count was 10.

The final result was generated by picking the lowest ranking item from each cluster, appending those to the result list, then repeating this until all items are used.

2.2 Run2: Text based reranking

The second subtask was the text based reranking which is accomplished using the title, tags and description fields of each image and then comparison based on TF-IDF numerical statistic and our distance calculating method.

As a preprocessing step we executed a stop word filtering and also remove some special characters (namely: .,:;0123456789()_@) and some remaining HTML parts (& and " and everything between < and >), then we used the remaining text as the input for a simple TF-IDF calculation [8].

Then we calculated the distance between images (e.g. description fields) A and B in the following manner. We initialize distance D as zero and compare A and B in term level. All occurring t terms in document A compared with all terms in the document B and so on. If term t is contained by both document, then $D2$ will be increased by 0. If t contained by only one document, we take into consideration the DF : if $DF < 5$, then it is a rare term and D increased by 2; if $DF > DN/4$, then it is a common term and $D2$ increased by 0.1 (where DN is the total number of documents). If the term is not common nor rare, then we added the $DF/documentcount$ to the distance.

Using the three text descriptors we created a weighted sum for the field distances, where the weight are as follows: title=1, tags=2, description=0.5

After we had the distance matrix we used the same reranking steps as in run1.

2.3 Run3: Text + Visual

In the third subtask both visual and textual descriptors could be used to create the results.

We used our visual distance matrix $D1$ and text distance matrix $D2$ and create a new aggregate matrix $D3$. This matrix is simply the sum of the corresponding values from both $D1$ and $D2$ matrix. We tried different kind of weighing methods, but the pure matrices supplied the best results on the devset without any further modification.

2.4 Run4: Credibility based reranking

In the fourth run participants were provided with credibility descriptors detailed in [2].

Using the original result we filtered the images by users who had *faceProportion* more than 1.3 to create the same effect as we did with the *FACE* metric. With the purpose of increase the diversity we used the *locationSimilarity* metric, if this value exceed the threshold of 3.0 we excluded the images. Despite our simple approach we had great results on the devset.

3. RESULTS AND CONCLUSION

run name	P@20	CR@20	F1@20
Visual single	.7022	.3702	.4751
Visual multi	.7164	.3857	.4813
Text single	.6435	.3494	.4379
Text multi	.7021	.3813	.4748
Vistext single	.6732	.3563	.4554
Vistext multi	.6993	.3683	.4651
Cred single	.7014	.3589	.4651
Cred multi	.7150	.3498	.4479

Table 1: Average results of the approaches

Our results can be seen on Table 3. and the F1 metrics can be seen on Figure 1, we listed the single and multi concept based results separately.

As you can see the visual based results are the best from the runs, from the devset we found out, that the text based information for the images are in a lot of cases missing, or don't represent the images well. It's not uncommon that an author gives the same textual information to all of the images in a topic.

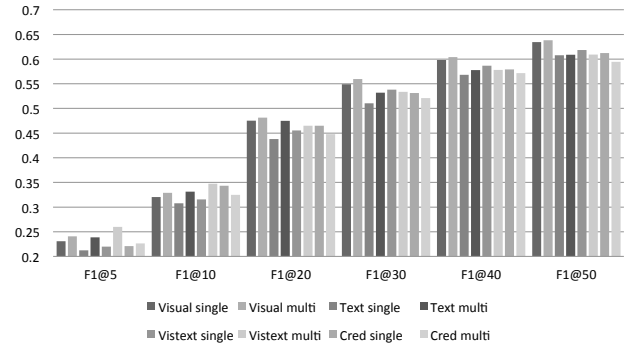


Figure 1: F1@N results

4. REFERENCES

- [1] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2):5:1–5:60, May 2008.
- [2] B. Ionescu, A. Popescu, A. Lupu, A. Ginsca, and H. Müller. Retrieving diverse social images at mediaeval 2015: Challenge, dataset and evaluation. In *Proceedings of the MediaEval 2015 Multimedia Benchmark Workshop*, 2015.
- [3] X. Ma, W. Wan, and L. Jiao. Spectral clustering ensemble for image segmentation. In *Proceedings of the First ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, GEC '09, pages 415–420, New York, NY, USA, 2009. ACM.
- [4] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, pages 849–856. MIT Press, 2001.
- [5] M. L. Paramita, M. Sanderson, and P. Clough. Diversity in photo retrieval: Overview of the imageclefphoto task 2009. In *Proceedings of the 10th International Conference on Cross-language Evaluation Forum: Multimedia Experiments*, CLEF'09, pages 45–59, Berlin, Heidelberg, 2010. Springer-Verlag.
- [6] Z. Paróczyi, B. Fodor, and G. Szucs. Dclab at mediaeval2014 search and hyperlinking task. In *Working Notes Proceedings of the MediaEval 2014 Workshop, Barcelona, Spain, October 16-17, CEUR-WS. org, ISSN 1613-0073*, 2014.
- [7] G. Szűcs, Z. Paróczyi, and D. Vincz. Bmemtm at mediaeval 2013 retrieving diverse social images task: Analysis of text and visual information. In *Working Notes Proceedings of the MediaEval 2013 Workshop, Barcelona, Spain, October 18-19, CEUR-WS. org, ISSN 1613-0073*, 2013.
- [8] J.-B. Yeh and C.-H. Wu. Video news retrieval incorporating relevant terms based on distribution of document frequency. In *Proceedings of the 9th Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing*, PCM '08, pages 583–592, Berlin, Heidelberg, 2008. Springer-Verlag.