

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 to the MediaEval 2015 Retrieving Diverse Social Images Task in order to increase the accuracy of the searching result of Flickr based on the metrics relevance and diversity. In our approach the new ranking is the modification of the original queue, considering and recalculating the diversity to every re-ranked item. We have used color related visual features and text descriptors to define similarity between images with our distance based spectral clustering method.

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) [1] 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 [4].

Our team participated in previous challenges [3, 2], 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.

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. RUNS

2.1 Run1: Visual based reranking

The main idea based on the previous papers [3, 2] is detecting and filtering photos with people (faces) on it, and after the preprocessing phase we build our distance matrix in order to define relevant and diverse images.

At first sight we designed our method using HOG features, but no significant results could be reached with HOG vector difference. Therefore we recommend a color only based clustering method, without using any other potential features such as shape or texture related descriptors.

The first preprocessing step is calculating two values to the image: first the *FACE* feature, which is the ratio of the faces on the image, and the second is the ratio of the black color ($CN[0]$) on each image. Using this two values on each image, we are able to increase the diversity and relevance of the result set by putting images to the end of the queue 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 $D1$ matrix between each A and B images: $D1_{AB} = |CN_A - CN_B| + [5 * Diff(CM_A, CM_B)]$, where $Diff$ function is the difference of the CM vectors with unique weighting on the different dimensions: $CM[0], ..., CM[2] : 1; CM[3], ..., CM[5] : 0.3; CM[6], ..., CM[8] : 0.1$.

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 execute a stop words filtering and also remove some special characters which have no value (e.g. html tags). At the next step we calculate TF-IDF values and distance between texts (e.g. description fields) A and B in the following manner. We initialize distance $D2$ as zero and compare A and B in term level. All occurring t terms in document A compared with all terms in document B and so on. If term t is contained by both document, then $D2$ will be increased by 0. Else, 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).

2.3 Run3: Text + Visual

We used our visual distance matrix $D1$ and text distance matrix $D2$ to create a new aggregate matrix $D3$. It is produced by simple summation of the corresponding values of both $D1$ and $D2$ matrix, where the actual item of each matrix corresponds to the same image. We tried different kind of weithing methods, but the pure matrices supplied the best results without any further modification.

2.4 Run4: Credibility based reranking

We present our solution to the credibility subtask which was refined and filtered by *faceProportion* and *locationSimilarity* values.

As a first step we re-ranked images which were originally at the end of the queue: each image after the 100th position in the original queue were put the end of the re-ranked queue, in order to increase the relevance of the result set. The *faceProportion* variable is used to purify the result set by putting the images to the end of the queue, where the face ratio is more than 1.3. With the purpose of increase the diversity we defined *locationSimilarity* with a threshold at 3.0. If the value of the variable exceed the threshold, the image will be moved to the end of the queue. This means, images with similar location will not be chosen to the result set repeatedly.

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

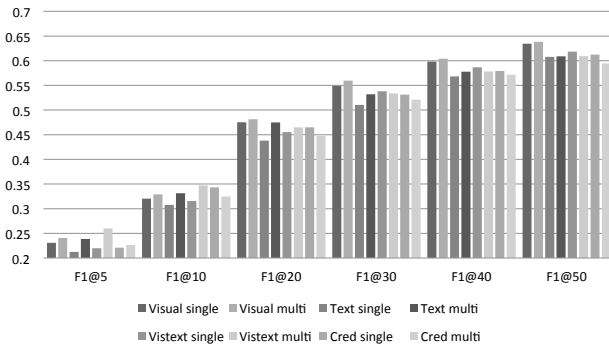


Figure 1: F1@N results

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

As you can see the visual based results are the best among the results, 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.

The credibility based results are better, then we expected, since our method basicly eliminates some authors from the results, which can have a negative effect on the CR metric.

4. REFERENCES

- [1] 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.
- [2] Z. Paróczy, 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.
- [3] G. Szűcs, Z. Paróczy, 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.
- [4] B. Taneva, M. Kacimi, and G. Weikum. Gathering and ranking photos of named entities with high precision, high recall, and diversity. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM '10*, pages 431–440, New York, NY, USA, 2010. ACM.