DCLab at MediaEval2014 Retrieving Diverse Social Images Task

Zsombor Paróczi
Inter-University Centre for Telecommunications and Informatics
paroczi@tmit.bme.hu

Bálint Fodor
Inter-University Centre for Telecommunications and Informatics
balint.fodor@aut.bme.hu

Gábor Szűcs
Inter-University Centre for Telecommunications and Informatics
szucs@tmit.bme.hu

ABSTRACT

TODO abstract

1. INTRODUCTION

Many potential tourists search on Web tries to find more information about a place he (or she) is potentially visiting. These persons have only a vague idea about the location, knowing the name of the place. Our aim is to help with these persons by providing a set of photos, as summary of the different views of the location. In the official challenge (Retrieving Diverse Social Images at MediaEval 2014: Challenge, Dataset and Evaluation) [1] a ranked list of location photos retrieved from Flickr (using text information) is given, and the task is , to refine the results by providing a set of images that are in the same time 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 refinement and diversification process can be based on the social metadata associated with the collected photos in the data set [2] and/or on the visual characteristics of the images. The initial results are typically noisy and redundant because of social media platform [3], where the large variety comes from very different users. The goodness of the refinement process can be measured by precision and diversity [6]. Earlier we have solved a very similar problem by diversification of initial results using clustering [5], but our solution was focused on only diversification. The largest development of this paper is that both of relevance and diversity are in the centre.

2. REORDERING SYSTEM

We took five approaches to generate the final reordering of the inital search result. This required five different systems that share similar components. All the systems take the inital ordering as the input along with the visual feature descriptors and the textual descriptors corresponding to the images. In every case the relevancy of each image is estimated, the images are grouped into clusters and based on this two type of information the final ordering is determined.

2.1 Relevance Scoring

run name	relevance	
run1	devset avg	
run2	devset avg	t
run3	devset avg	visua
run4	devset avg scaled with user credibility	visua
run5	combined devset avg and credibility information	visua

Table 1: TODO caption

For every kth place in the orderings of the developer data set we calculated the probability of the item at the kth place is beeing relevant. Before giving the formal definition let denote the set of all orderings in the developer set as L, the kth element of the ordering $l \in L$ as l_k and the binary function of the relevancy (based on the ground truth data) as $r_{gt}(l_k)$. Then p_k, the estimated probability of the kth element in an ordering is relevant: $p_k = \frac{1}{|L|} \sum_{l \in L} r_{gt}(l_k)$.

When processing an ordering (from the test data set) we give the relevance score of p_k to the kth element of the ordering.

2.2 Clustering

TODO text clustering

The provided data sets contain visual feature descriptors (color moments, histogram of oriented gradients, etc.) in csv files. First, we merged the descriptors into a long feature vector, one vector for an image. The clustering is done on a per ordering basis, so the feature vector is calculated for every image in an ordering. Then the components of the vectors are normalized to bring all the data to the same scale. The vectors are clustered with the K-means algorithm trying all the number of clusters parameter from 6 to 18. For every clustering the silhouette score [4] is calculated and the best instance is selected.

Clusterings based on the textual and the visual data can differ, but merging the two results can be beneficial. Having two clustering functions $c_1(x)$ and $c_2(x)$ that are mapping an image id to a cluster label, one can construct $c_3(x) = (c_1(x), c_2(x))$ that maps an image id to a new cluster labeled by the pair of the two original cluster labels. Note that the new label set is the Cartesian product of the two original cluster label sets.

2.3 Final Ordering

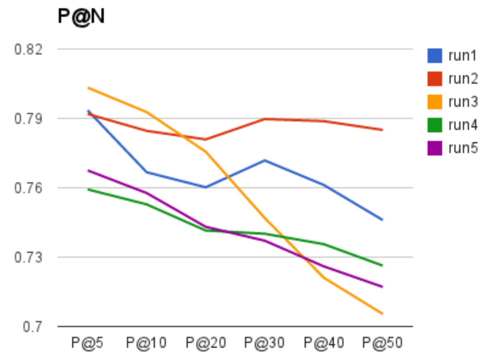
TODO

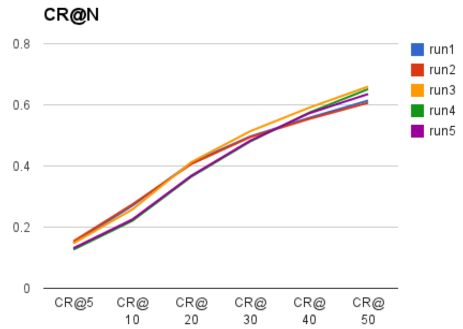
3. RESULTS AND CONCLUSION

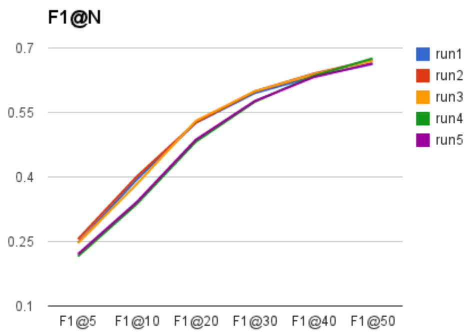
MediaEval 2014 Workshop, October 16-17, 2014, Barcelona, Spain

run name	P@20	CR@20	F1@20
VisClusterAvgRelevance	.7602	.4107	.5259
TextClusterAvgRelevance	.7809	.4065	.527
VisTextClusterAvgRelevance	.7756	.4127	.5305
VisTextClusterCredRelevance	.7415	.3651	.4819
VisTextClusterMixedRelevance	.7431	.3682	.4866

Table 2: Average results of the five approaches







4. ACKNOWLEDGMENTS

The publication was supported by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

5. REFERENCES

[1] B. Ionescu, A. Popescu, A. Lupu, A. Ginsca, and H. Müller. Retrieving diverse social images at mediaeval 2014: Challenge, dataset and evaluation. In *Proceedings of the MediaEval 2014 Multimedia Benchmark Workshop*, 2014.

[2] B. Ionescu, A.-L. Radu, M. Menéndez, H. Müller, A. Popescu, and B. Loni. Div400: a social image retrieval result diversification dataset. In *Proceedings of the 5th ACM Multimedia Systems Conference*, pages 29–34. ACM, 2014.

[3] A.-L. Radu, B. Ionescu, M. Menéndez, J. Stöttinger, F. Giunchiglia, and A. De Angeli. A hybrid machine-crowd approach to photo retrieval result diversification. In *Proceedings of International Conference on MultiMedia Modeling, LNCS 8325*, pages 25–36. Springer, 2014.

[4] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[5] G. Szűcs, Z. Paróczi, 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.

[6] 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.