Reordering the image search results for relevance and diversity in MediaEval 2014 Challenge

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 do websearches when they try to find more information about a place he (or she) is potentially visiting. These people have only a vague idea about the location, knowing the name of the place. Our aim is to help them 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 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 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 using the precision and diversity metric [6]. Earlier we have solved a very similar problem by diversification of initial results using clustering [5], but our solution was focused on diversification only. The largest development of this paper is that we focus on both the relevance and diversity.

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 each case the relevancy of all images are estimated, the images are grouped into clusters and based on this two type of information the final ordering is determined.

Table 1. shows the different system compositions we used. Section 2.1. describes the 'average' ('avg') relevance estimation and its extended versions utilizing user credibility information. User credibilities are used with different weights in the last two approaches. Section 2.2. defines the methods we used to cluster the data.

run name	relevance	clustering
run1	avg	visual
run2	avg	textual
run3	avg	visual+textual
run4	avg + credibility 1	visual+textual
run5	avg + credibility 2	visual+textual

Table 1: Reordering approaches.

2.1 Relevance Scoring

For every kth place in the orderings of the developer data set we calculated the probability of the item at the kth place is 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.

In table 1. 'avg + credibility' means that the relevance estimation is multiplied by the user credibility (in range $[0, 1]$).

2.2 Clustering

The provided data sets contains 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 each image. The clustering is done per ordering, 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 by trying all clustering number parameters from 6 to 18. For every clustering the silhouette score [4] is calculated and the best instance is selected.

Clusterings based on textual and 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

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

cluster label sets.

2.3 Final Ordering

Our reordering algorithm (in order to get maximal F_1 value in each subset of the answer list) consists of four phases.

1) Take the elements in each cluster in descending order and select the element that possessing the largest probabilities of relevance, this will be the 1st in the reordered list. 2) Lth step: take the first elements in each cluster as candidate and calculate the estimated F_1 measure:

$$F_1(@L) = \frac{2 \cdot P(@L) \cdot CR(@L)}{P(@L) + CR(@L)}$$

3) Select the element possessing the largest estimated F_1 measure and move to the reordered list. 4) Continue with phase 2 until we have cluster elements left.

3. RESULTS AND CONCLUSION

Figures 1., 2., 3. shows the values of P@N, CR@N and F1@N for the different runs, while table 2. shows the average P@20, CR@20 and F1@20 results.

run name	P@20	CR@20	F1@20
run1	.7602	.4107	.5259
run2	.7809	.4065	.527
run3	.7756	.4127	.5305
run4	.7415	.3651	.4819
run5	.7431	.3682	.4866

Table 2: Average results of the five approaches

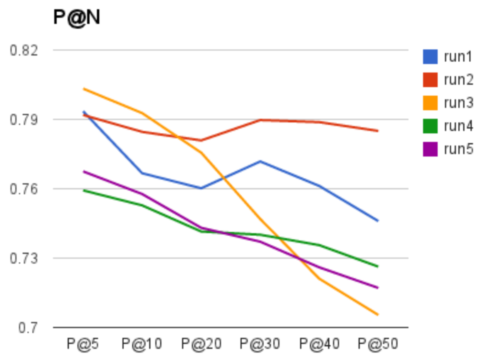

Figure 1: P@N results

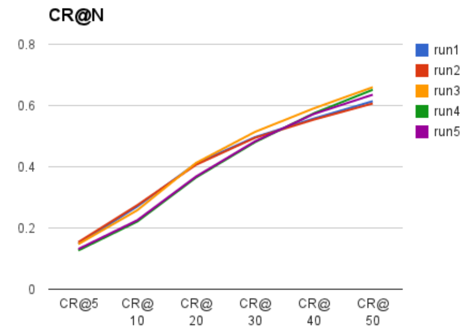

Figure 2: CR@N results

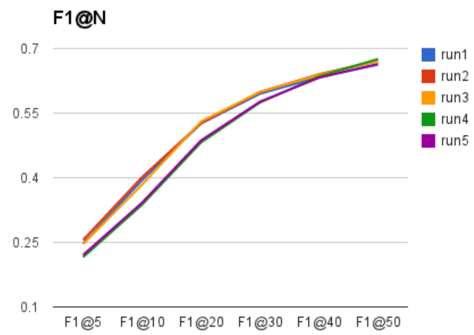

Figure 3: F1@N results

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.