

# Re-ranking the image search results for relevance and diversity in MediaEval 2014 Challenge

Zsombor Paróczy  
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

In this paper we introduce a refinement and diversification process for re-ranking image search results based on social metadata and visual characteristics of the photos. The goal of the developed re-ranking algorithm is to construct a new sequence with maximal value of the harmonic mean of precision and diversity. Our contribution is twofold: estimation of precision using the statistical average and mixing of clustering results in order to get better diversity. In the combined clustering the new label set is the Cartesian product of the two original cluster label sets.

## 1. INTRODUCTION

Many potential tourists do websearches when they try to find more information about a place they are 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 of the MediaEval 2014 Retrieving Diverse Social Images task [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 the imperfect metadata and the current, restricted search capabilities of the social media platforms [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. In a previous participation in the task we have solved a very similar problem via diversification of initial results using clustering [5], but our solution was focused on diversification only. In this paper we focused on relevance and diversity with the same importance, as a new idea.

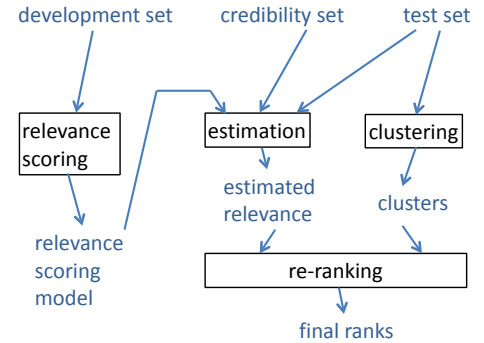


Figure 1: System overview

## 2. RE-RANKING SYSTEM

We took five approaches to generate the final re-ranking of the initial search result. This required five different systems that share similar components. The components can be seen in Figure 1, where the 'relevance scoring' part creates a model, based on this model the 'estimation' component estimates the relevance of test data (optionally using the credibility set), 'clustering' part generates clusters, and 're-ranking' component produces final ranks.

All the systems take the initial ranks as input along with the visual feature descriptors and the textual descriptors corresponding to the images. In each case the relevance of all images are estimated, the images are grouped into clusters and based on this two type of information the final ranks are determined.

Section 2.1. describes the 'average' relevance estimation and its extended versions utilizing user tagging credibility information. Tagging credibilities are used with different weights in the last two approaches. Section 2.2. defines the methods we used to cluster the data.

### 2.1 Relevance Scoring and Estimation

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

Table 1: Re-ranking approaches.

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

When processing an initial sequence (from the test data set) we give the relevance score of  $p_k$  to the  $k$ th element of the sequence. In Table 1 'avg + credibility' means that the relevance estimation is multiplied by the user tagging credibility.

## 2.2 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 each image. 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 cluster label sets.

## 2.3 Final Re-ranking

Our re-ranking 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 probability of relevance, this will be the 1st in the reordered list. 2)  $L$ th step: take the first elements in each cluster as candidate and calculate the estimated  $F_1$  measure:

$$F_1(@L) = \frac{2 \cdot P(@L) \cdot CR(@L)}{P(@L) + CR(@L)} \quad (1)$$

3) Select the element that has the largest estimated  $F_1$  measure and move to the re-ranked list. 4) Continue with phase 2 until we have cluster elements left.

## 3. RESULTS

Table 1 shows the different system compositions we used. 'Avg' is for the 'average' relevance estimation detailed in section 2.1. Figure 2 shows the values of  $F_1$  for the different runs, while Table 2 shows the average P@20, CR@20 and F1@20 results. The results of the five approaches mainly differ in P@N performance. The cluster recall (CR) is almost the same, so P@N has more impact on the  $F_1$  score.

The credibility information tends to have negative effect on both P and CR in our tests, compared to the other runs.

However run3 (clustering based on visual+textual data) underperforms run2 in P@20 average, it is more diverse, so the overall F1@20 score is the highest for run3.

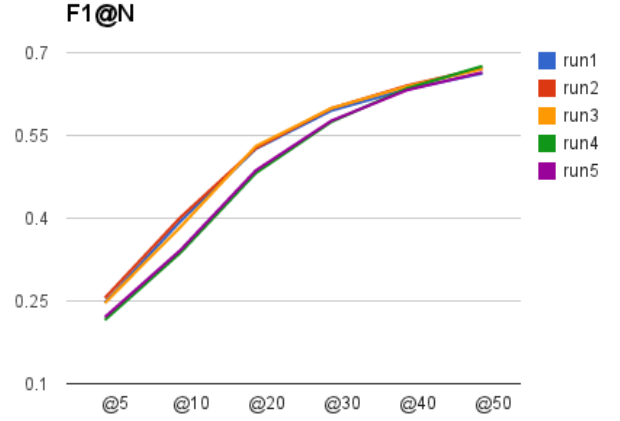


Figure 2: F1@N results

Table 2: Average results of the five approaches

run name	P@20	CR@20	F1@20
run1	0.7602	0.4107	0.5259
run2	0.7809	0.4065	0.527
run3	0.7756	0.4127	0.5305
run4	0.7415	0.3651	0.4819
run5	0.7431	0.3682	0.4866

## 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.