

TIA-INAOE's approach for the 2013 Retrieving Diverse Social Images task*

Hugo Jair Escalante, Alicia Morales-Reyes
Instituto Nacional de Astrofísica, Óptica y Electrónica
Luis Enrique Erro 1, 72840, Puebla, Mexico
{hugojair, a.morales}@inaoep.mx

ABSTRACT

This paper describes the approach adopted by the TIA-INAOE team for the 2013 Retrieving Diverse Social Images task of MediaEval. The challenge consists in re-ranking a list of images returned by a retrieval system in such a way that visual diversity among images in the first positions is maximized [5]. A database of partially-annotated images of tourist destinations is provided. The problem is tackled as an optimization one with the aim of finding a new image ranking that shows improved diversity according to the criterion defined by MediaEval. The proposal is to apply a multi-objective evolutionary algorithm to simultaneously maximize image diversity in consecutive positions of the list and minimize divergence from the original list. Multi-modal information can be incorporated by the proposed approach. Results obtained in the MediaEval forum are reported and analyzed.

Keywords

Evolutionary algorithms; Multi-objective optimization; Retrieval result diversification; Multi-modal image retrieval

1. INTRODUCTION

Retrieval results diversification has been a very active research-topic and recently has had an increasing popularity in information retrieval. In particular, diversification is essential when searching very large image collections. For instance, users do not want to see the same or very similar views/images (e.g., the Eiffel tower) regarding a particular query/topic (e.g., Paris), even when those images are relevant to the query. Instead, it is desirable that multiple views associated to the topic are shown in the first results (e.g., Eiffel Tower, Arch of Triumph, Notre Dame, etc.). Therefore, maximizing relevancy should not be a unique objective for image retrieval systems, but a trade-off between relevance and diversity must be aimed.

In this note, the solution to the 2013 Retrieving Diverse Social Images MediaEval Task proposed by the TIA-INAOE¹ research group is outlined. A detailed description of the considered scenario is provided in [5]. In a nutshell, the goal is

*This work was partially supported by the LACCIR program under project id R1212LAC006.

¹<http://ccc.inaoep.mx/~tia/>

to develop a method able to re-rank a list of images, returned by an image retrieval system, in such a way that images in the first ranking positions are visually diverse to each other. In addition, the images in the list must be ranked in descending relevance order according to the query. Hence, the proposed approach targets diversification as an optimization challenge of two objectives: relevance and diversity [3].

2. RELATED WORK

Several approaches for result diversification have been proposed: Arni et al. reported results obtained by several strategies evaluated in the ImageCLEF2008 photographic retrieval task which focused on diversification [1]. Clustering, topic modeling and margin-maximization approaches were proposed to re-rank images. Whereas these methods proved to be effective, they still attempt to optimize a single criterion. Deselaers et al. sustained that diversification involves two conflicting objectives: relevance and diversity [3]. Their proposal is an effective dynamic programming approach to optimize an objective function that combines relevance and diversity estimates into a single term. This paper proposes a multi-objective evolutionary algorithm that explicitly attempts to optimize relevance and diversity objectives.

3. PROPOSED METHOD

Considering that a list of N images ($L = \langle I_1, \dots, I_N \rangle$) relevant to a particular query (Q) has an associated ranking score: $S^0 = \langle s_1, \dots, s_N \rangle$, where $s_i \geq s_j, \forall i, j : i \neq j$ and $i < j$ (if the scoring value is unknown, an estimation is calculated by $s_i = \frac{1}{i}$). The goal is to find the ranking score $S^* = \langle s_1^*, \dots, s_N^* \rangle$ such that the ranking induced by S^* maximizes the objectives of Equations (1) and (2):

$$\rho(S^0, S) = 1 - \frac{6}{n(n^2 - 1)} \sum_i dr_i(S^0, S)^2 \quad (1)$$

where $dr_i(S^0, S)$ is the difference in rankings at position i induced by scores lists S^0 and S . Spearman's correlation coefficient (ρ) measures discrepancy between S and the initial scores in S^0 . Implicitly, it is assumed that the initial list is a good one in terms of relevance.

Equation (2) is defined to evaluate visual similarity among images ranked by S in consecutive positions:

$$\beta(S) = \sum_{i=2}^N \min(d_d(I_i, I_{1, \dots, i-1})) \quad (2)$$

Table 1: Summary of submitted runs.

ID	Description
1	MORD - HOG / Euclidean distance for d_d
2	MORD - Tags + descriptions / cosine dissim. for d_r
3	MORD - 3 objectives: ρ , visual d_d^v and textual d_d^t div.

where β is the diversity term, $d_d(I_i, I_{1,\dots,i-1})$ measures the visual distance between image in the i^{th} rank position and the rest of the images appearing in previous positions. Since $d_d(I_i, I_{1,\dots,i-1})$ is not associated to a particular feature representation, it can be estimated by using any visual feature provided for the task [5]. Moreover, d_d can be estimated by using textual information or meta-data associated to the images. Calculating β in this way to represent diversity modifies the diversity term defined in [3].

Aiming to find the score list S^* that offers the best trade-off between ρ and β is the main objective of this study. This problem can be tackled by a multi-objective evolutionary optimization technique maximizing simultaneously ρ and β . It was decided to use the NSGA-II algorithm to target this goal [2]. NSGA-II is one of the most used multi-objective evolutionary algorithms. Standard operators for selection, crossover and mutation were adopted. The NSGA-II algorithm returns as output the set of non-dominated solutions (i.e., a set of re-ranked lists that optimize ρ and β), an estimate of the Pareto front for the problem at hand. Theoretically, none of these solutions is better or worse than the others, therefore all of them are valid solutions. However, for our problem, a single solution has to be selected, thus a strategy for selecting a single solution from the set of solutions is also proposed. Specifically, values normalization of the involved objectives is carried out across the returned solutions and the sum of normalized objectives is ranked. In this way, the solution at the first position offers a good trade-off between relevance and diversity.

4. EXPERIMENTAL RESULTS

Three runs of the proposed Multi-Objective Result Diversification (MORD) method were submitted to be considered for evaluation, these are summarized in Table 1. In a visual run, HOG features are used to estimate d_d , this choice is justified by preliminary experimentation in the development data set. Also, a textual run was submitted in which the term d_d was estimated by the cosine dissimilarity between the bag-of-words (BoW) representation obtained by tags and the textual images descriptions. For the textual information character 3-grams were used instead of words for building the BoW. This type of representation is particularly helpful for texts dealing with informal language, because writing style patterns can be discovered (see e.g., [4]). Finally, a multi-modal run was also submitted where three objectives are simultaneously maximized: ρ and diversity terms d_d^v and d_d^t , considered for visual and textual runs respectively.

The average (test-set) performance for a number of measures and for the three runs are reported in Table 2. We show results using expert and crowd-sourcing ground-truth (a 50 images sample was evaluated and the average over three subjects is reported). Regarding expert’s evaluation, it can be seen that performance difference among three runs is roughly the same. Slightly better performance was ob-

Table 2: Official results obtained by MORD.

Expert evaluation						
	C@10	C@20	P@10	P@20	F@10	F@20
1	0.3808	0.5699	0.7146	0.7143	0.4795	0.6125
2	0.3885	0.5732	0.7091	0.7136	0.4801	0.6102
3	0.3823	0.5690	0.6977	0.7076	0.4728	0.6067
Crowd-sourcing evaluation						
	C@10	C@20	P@10	P@20	F@10	F@20
1	0.7194	0.8499	0.6755	0.6745	0.6640	0.7245
2	0.7503	0.8625	0.6755	0.6898	0.6790	0.7392
3	0.7479	0.8675	0.6714	0.6918	0.6769	0.7464

tained when using only textual data to estimate diversity among images (run 2). This is a somewhat unexpected result because the aim is to maximize visual diversity. Regarding the crowd-sourcing evaluation, a similar pattern is observed, although the results are much higher than in the expert evaluation. It is difficult to determine how good these results are on test data since other systems results have not been revealed yet. However, from results reported in this working note, it is possible to observe that the achieved performance on test data is similar regardless of the modality for the proposed method. Nevertheless, it is worth mentioning that during the development phase, performance of the proposed method was evaluated and seemed competitive. For instance, the multi-modal run using development data (*keywords* only) obtained a C@R10 of 0.5043 ± 0.0111 (10 runs of MORD) compared to 0.4635, which is the C@R10 obtained when using the base system (using the list induced by S^0), this is a relative difference of $\approx 8.9\%$.

5. DISCUSSION AND FUTURE WORK

In this study, visual results diversification has been tackled as a multi-objective optimization problem while considering relevance and diversity as two independent but complementary objectives. To the best of our knowledge, this is the first approach to explicitly optimize both objectives by using a multi-objective evolutionary optimization technique.

6. REFERENCES

- [1] T. Arni, P. Clough, M. Sanderson, and M. Grubinger. Overview of the imagecelephoto 2008 photographic retrieval task. In *CLEF 2008*, volume 5706 of *LNCS*, pages 500–511. Springer, 2009.
- [2] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.
- [3] T. Deselaers, T. Gass, P. Dreuw, and H. Ney. Jointly optimising relevance and diversity in image retrieval. In *Proc. of ACM Conference on Image and Video Retrieval*, pages 1–8, 2009.
- [4] H. J. Escalante, T. Solorio, and M. Montes y Gomez. Local histograms of character n-grams for authorship attribution. In *Proc. of ACL*, pages 288–298, 2011.
- [5] B. Ionescu, M. Menéndez, H. Müller, and A. Popescu. Retrieving diverse social images at mediaeval 2013: Objectives, dataset and evaluation. In *MediaEval 2013 Workshop*, Barcelona, Spain, October 18–19 2013.