

Ghent University-iMinds at MediaEval 2013 Diverse Images: Hierarchical Clustering...

Baptist Vandersmissen¹
baptist.vandersmissen@ugent.be

Abineshwar Tomar¹
abineshwar.tomar@ugent.be

Frédéric Godin¹
frederic.godin@ugent.be

Wesley De Neve^{1,2}
wesley.deneve@ugent.be

Rik Van de Walle¹
rik.vandewalle@ugent.be

¹ ELIS - Multimedia Lab, Ghent University – iMinds, Ghent, Belgium

² Image and Video Systems Lab, KAIST, Daejeon, South Korea

ABSTRACT

In this paper, we attempt to tackle the MediaEval 2013 Retrieving Diverse Social Images challenge, which is a filter and refinement problem on a Flickr based ranked set of social photos...

1. INTRODUCTION

In this paper, we describe our approach for tackling the MediaEval 2013 Retrieving Diverse Social Images Task [1]. This task focuses on result diversification in the context of social photo retrieval. A modern retrieval system focuses almost exclusively on the accuracy of the results related to a certain information need. These retrieval systems often result in very relevant but also very similar photos. This task aims at improving the retrieval technology linked to social photos with the goal of retrieving not only representative but also diverse photos depicting the information need in a complete manner.

Photos of popular and less popular monuments are fetched from Flickr¹ (based on name and/or GPS coordinates) and ranked with its default “relevance” algorithm. Apart from a list of photos, a large set of visual and textual features derived from these photos are given. The dataset is subdivided into a development set and a test set and contains a total of 396 monuments. The requirements of this task are to provide three runs on the test set, respectively focusing on solely visual features, textual features and a combination of both to perform a refinement of the Flickr ranking. This refinement is a list of maximum 50 newly ranked photos that are considered both relevant and diverse representations of the information need.

We suggest a cluster based approach for the visual run and a textual run based on semantic similarity. The third run focuses on hierarchical clustering of relevant photos and represents a combination of the purely visual and textual techniques. In the next sections, we will discuss these approaches in more detail.

¹www.flickr.com

2. VISUAL RUN

3. TEXTUAL RUN

The textual run solely makes use of information derived from tags and other textual metadata. This approach aims to diversify the results by reranking the images retrieved from Flickr based on textual relevance and semantic similarity. Our solution is based on [2] and makes use of an adapted performance metric that takes into account both relevance and diversity to evaluate a ranking. Images for a query can then be ordered by directly optimizing the performance metric. This metric is named Average Diverse Precision (ADP) and is derived from the conventional Average Precision metric by adding a diversity component. We refer to [2] for a comprehensive overview.

Optimizing the expected value of the ADP measurement is a permutation problem of complexity $O(n!)$. Therefore a greedy approach that optimizes an estimation of the ADP measurement is implemented. Denote by τ an ordering of the images, and let $\tau(i)$ be the image at the position of rank i (a lower number indicates image with a higher rank). With the top $i - 1$ documents established, we can derive that the i th image should be decided as follows

$$\tau(i) = \arg \max_{x \in \mathcal{D} - \mathcal{S}} \frac{Rel(x)}{i} Div(x)(C + Div(x)), \quad (1)$$

where

$$\mathcal{S} = \{\tau(1), \tau(2), \dots, \tau(i - 1)\}, \quad (2)$$

$$C = \sum_{k=1}^{i-1} Rel(\tau(k)) Div(\tau(k)). \quad (3)$$

Where $Rel(x)$ and $Div(x)$ respectively denote the estimated relevance and diversity of the image.

3.1 Relevance Estimation

We try to measure the relevancy of an image related to a certain monument. In order to achieve this we make use of the textual metadata such as number of views, number of comments and social tags. The task provides this data together with three textual features: TF-IDF, Social TF-IDF and probabilistic model. We suggest a linear combination of all gathered data.

$$Rel(x) = \alpha \times tags_x + \beta \times views_x + \gamma \times comments_x,$$

where

$$tags_x = \frac{1}{|\mathcal{T}_x|} \left(\sum_{t \in \mathcal{T}_x} a \times prob_{t,x} + b \times tfidf_{t,x} + c \times stfidf_{t,x} \right),$$

is also a linear combination of above described features. We found that the precision of image orderings, which are purely based on relevance information, can be maximized by setting parameters a and γ to be zero and increasing the influence of tags and views. Thus, implicating that the number of comments and probabilistic model information do not add extra information to the relevance estimation of an image.

3.2 Diversity Estimation

Diversity of an image is defined as the minimal difference with the other images in the ranking.

$$Div(x) = \min_{i \in \{1, \dots, n\}} (1 - s(x, \tau(i)))$$

While difference between two images is defined by subtracting a similarity measure from one. We take into account

4. VISUAL AND TEXTUAL RUN

5. EXPERIMENTS

6. CONCLUSIONS

7. REFERENCES

- [1] M. Bowman, S. K. Debray, and L. L. Peterson.
Reasoning about naming systems. *ACM Trans.
Program. Lang. Syst.*, 15(5):795–825, November 1993.
- [2] M. Wang, K. Yang, X.-S. Hua, and H.-J. Zhang.
Towards a relevant and diverse search of social images.
Multimedia, IEEE Transactions on, 12(8):829–842,
2010.