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

Baptist Vandersmissen[1] baptist.vandersmissen@ugent.be
Abineshwar Tomar[1] abineshwar.tomar@ugent.be
Fréderic Godin[1] frederic.godin@ugent.be
Wesley De Neve[1,2] wesley.deneve@ugent.be
Rik Van de Walle[1] 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 images...

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 images. This task aims at improving the retrieval technology linked to social images with the goal of retrieving not only representative but also diverse images depicting the information need in a complete manner.

Photos of popular and less popular monuments are fetched from Flickr[1] (based on name and/or GPS coordinates) and ranked with its default "relevance" algorithm. Apart from a list of images, a large set of visual and textual features derived from these images 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 images 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 images and represents a combination of the purely visual and textual techniques. In the next sections, we will discuss these approaches in more detail.

[1]www.flickr.com

MediaEval 2013 Workshop, October 18-19, 2013, Barcelona, Spain

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 measurment is a permutation problem of complexity $O(n!)$. Therefore a greedy approach that optimizes an estimation of the ADP measurement is implemented. Denote by $\boldsymbol{\tau}$ 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 ith image should be decided as follows

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

where

$$\mathscr{S} = \{\tau(1), \tau(2), \ldots, \tau(i-1)\}, \quad (2)$$

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

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 and normalized data.

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

Table 1: Results on development set (comparison with Flickr) and test set.

	Development Set (comparison with Flickr)				Test Set			
	+ GPS		- GPS		+ GPS		- GPS	
	CR@10	P@10	CR@10	P@10	CR@10	P@10	CR@10	P@10
Visual	43.6 (2.4)	79.2 (-6.8)	48.4 (2.0)	71.6 (2.8)	37.5	76.1	34.7	56.8
Textual	44.2 (2.9)	81.6 (-4.4)	51.6 (5.2)	67.2 (-1.6)	39.7	74.9	37.5	58.6
Combined	**49.8 (8.5)**	**85.6 (-0.4)**	**51.7 (5.3)**	**74.8 (6.0)**	**41.3**	**80.5**	**42.8**	**66.7**

where

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

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 andγ 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.

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

We use a semantic similarity measure based on google distance [**?**] to assess the difference between two images. The average of the summation of the similarities between the different tags of both images gives us a value that describes the semantic similarity between both images.

4. VISUAL AND TEXTUAL RUN

With the availability of both textual and visual data we can improve above methods. We use textual data to estimate the relevancy of an image while visual data is used for assessing the similarity. Next, we give a brief overview of the implemented algorithm.

4.1 Overview

To estimate the relevancy of an image we use the textual based method (cf. Section 3.1). Similarity between images is based on visual image features and a gaussian kernel method to measure the difference between two image vectors (cf. Section 2). In order to provide both a relevant and diverse ordering we make use of hierarchical clustering techniques. Assume we want to refine a set of m images retrieved from flickr to a ranking of size n.

- First, k images are selected with the highest estimated relevancy. k is an arbitrary number that depicts a subset of the final ranking. The larger k becomes the more the focus will shift from relevancy to diversity. Our algorithm sets $k = 10$. Denote that k is a measure for the number of images that is clustered and optimalized for diversity purposes.
- Next, these n images are hierarchically clustered based on a distance matrix calculated with above described visual similarity.
- Per cluster the most relevant item is selected and added to the final ranking.
- Depending on the number of remaining places of the first k spots in the final ranking images are greedily added based on a **gain** score. This score is higher for images that maximize the diversity and relevancy related to the current ranking.
- When the first k spots in the ranking are filled the algorithm starts from the beginning until all n spots are taken.

5. EXPERIMENTS

To evaluate our methods we evaluate the created rankings of monuments in the development set based on the delivered ground truth. We compare the measurements with the regular Flickr rankings to get a general idea of the functioning of our algorithm. In Table **??** we can see the results of our created algorithms (three runs) compared to regular Flickr ranking.

Table **??** shows the overall results of our algorithms on the test set. We notice the superior performance of the combined method over all other methods.

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.