UNED-UV@2016 Retrieving Diverse Social Images Task

A. Castellanos
NLP&IR Group UNED, Madrid
acastellanos@lsi.uned.es

X. Benavent
Informatics, U. Valencia
xaro.benavent@uv.es

A. García-Serrano
NLP&IR Group UNED, Madrid
agarcia@lsi.uned.es

E. de Ves
Informatics, U. Valencia
esther.deves@uv.es

ABSTRACT

This paper details the participation of the UNED-UV group at the MediaEval 2016 Retrieving Diverse Social Images Task using a multimodal approach. Several Local Logistic Regression models, which use the visual low-level features, estimate the relevance probability for all the images in the dataset. Then, the images are ranked by selecting the highest probability image at each of the textual clusters. These textual clusters are generated by making use of a textual algorithm based on Formal Concept Analysis (FCA) and Hierarchical Agglomerative Clustering (HAC) to detect the latent topics addressed. The images will be then diversified according to detected topics.

1. INTRODUCTION

Information retrieval systems have mainly relied on optimizing the result list according to accuracy-based metrics. However, when dealing with image retrieval, systems should be able to offer relevant but also diverse results [1]. Then, we propose a multimodal approach that works with relevance and diversity. It uses a relevance feedback algorithm developed by the UV group [3, 6, 7]. This method estimates the similarity probability of all the images from the dataset using visual low-level features by means of several Local Logistic Regression models (LLR) [10]. It has extensively been proven that the visual information has a great impact in the information retrieval systems [12]. For diversification, we propose an approach to represent an image by applying the concepts covered by the textual information of the images. This conceptual representation is tackled by means of Formal Concept Analysis (FCA) [13], a data organization technique. In our participation in previous editions of this task, we proved that this approach was able to identify the different topics addressed in the images, allowing the diversification of the results list according to them [5, 4].

2. SYSTEM DESCRIPTION

We present a two-step system that ranks a list of retrieved images taking into account the relevance of the retrieved images to the given query and showing as much diversity as possible. The first step is based on a relevance feedback algorithm that estimates the relevance of each image to the query by using the provided visual low-level features of the images [8]. In the next step, we select the highest probability image at each textual FCA cluster or visual k-means clustering to generate the list of similar and diverse images.

2.1 Relevancy via relevance feedback

We try to get the relevance of each of the images to the given query by using a relevance feedback algorithm [7]. The general methodology involves four steps:

1. *Reduction of the data dimensionality.* The provided low-level visual features [8] are used to generate a feature vector associated to each image that will be generically denoted as $\boldsymbol{x}$ in a dimensional space $N = 8194$. These features are reduced using a Principal Component Analysis (PCA). We retain only the first components that account for 80% of the data variability. We have used this idea to reduce the original dimension of our characteristic space in a new characteristic vector of dimension $M < N$, being $M = 52$. One of the advantages of this reduction is that the new transformed components are in decreasing order with respect to the variance explained by the corresponding principal component;

2. *Selecting the relevant and non-relevant sets.* The user looks at a few screens, each showing some images and marks some of them as being relevant and non-relevant (*run4*). For the automatic runs (*run1*, *run3* and *run5*), these sets are automatically selected. Relevant images are the first P images of the ranked Flicker list that belong to different textual-FCA o visual (k-means) clusters, being $P = 5$. The goal of selecting relevant images from different clusters is to give diverse relevant example images for the model to improve the diversity in the estimation. We use a K-means clustering by using provided visual low-level features in *run1* and *run5*, and a textual FCA cluster by using provided textual features [8] in *run3*. Non-relevant images are selected from other topics at each query;

3. *Parameter estimation of the Local Logistic Regression Models [10].* The reduced feature vectors (PCA) and the relevant and non-relevant sets are the inputs or several Local Logistic Regression models whose outputs are the probabilities that an image belongs to the relevant set. The feature vector is splitted dynamically in m groups of non-fixed size. Each group is used for adjusting the model of higher order, given the inputs sets and PCA components;

4. *Ranking of the database.* Models are evaluated on all the images of the database and return the probabilities of being relevant for each estimated model; as results, we have a probability vector ($\boldsymbol{p}$) of dimension m for each individual

MediaEval 2016 Workshop, October 20-21, 2016, Hilversum, Netherlands.

image. We combine these probabilities in just one by using a weighted average. The weights ($\boldsymbol{w}$) for a given probability are obtained by the amount of variance accounted for the group of components used to adjust the model. Finally, this procedure gives us a score/probability for each image.

2.2 Clusters for diversity

The final rank is generated by selecting the highest probability image from the different clusters trying to give as much diversity as possible to the diverse final list. If there are less than 50 clusters, a second highest probability image selection is done. We have generated clusters for the ranking using textual information, from *run2* to *run5*, and k-means clusters by using the visual information of the images (*run1*).

2.2.1 Textual clusters with FCA

The presented clustering procedure is based on the discovering of the latent topics addressed by the textual information in the images. To that end, Formal Concept Analysis is proposed to detect these topics and a Hierarchical Agglomerative Clustering (HAC) [11] to group similar images together into the detected topics. Each HAC-based cluster contains the image set covering a similar topic. Thereafter, the images of each cluster are ranked according to their diversity based on their visual features.

FCA-based Modelling.

Formal Concept Analysis (FCA) is a theory of concept formation [13] for the organization of content according to their related features. The basic formation of FCA is the *formal context*, a structure $\mathbb{K} := (G, M, I)$, where G is a set of objects, M a set of attributes related to these objects and I a binary relationship between G and M, denoted by gIm: the object g has the attribute m. From the *formal context*, a set of *formal concepts* can be inferred i.e., a *formal concept* is a pair (A, B) of images A and the features shared by those images B and organized in a *lattice* from the most generic to the most specific one. By applying FCA to the formal context containing the textual information of the images, they are modelled in terms of *formal concepts*, which group together the images sharing a same set of features. In order to select only those most-representative features, we applied Kullback-Leibler Divergence (KLD) [9] on the textual contents related to the images. This KLD-based selection represents each image by the textual contents that better differentiates a image from the other ones.

HAC-based grouping.

From the FCA *formal concepts*, a set of diverse image groups is created by applying a HAC algorithm [11]. Specifically, we propose a Single Linking Hierarchical Clustering that groups together similar *formal concepts* and the Zero-Induces index to set the cluster similarity [2].

2.2.2 Visual clusters

The clusters are made by a k-means procedure ($k = 18$) over the PCA components of the provided visual low-level features [8].

3. RESULTS

We submitted five runs (see table 1), four of them are automatic (*run1*, *run2*, *run3* and *run5*), and one, *run4* a human-based run. All of them use our two-step system except *run2* that uses only the second step, ranking the images by textual clusters with FCA. Three of them are multimodal runs (*run3*, *run4* and *run5*) using both textual and visual information.

Table 1: Description of the runs.

	Relevance algorithm					Ranking	
	Relevant images			Non-relevant images		clusters	
	text	visual	human	other	human	text	visual
run1		✓		✓			✓
run2						✓	
run3	✓			✓		✓	
run4			✓		✓	✓	
run5		✓		✓		✓	

Table 2: Official Metrics for Retrieving Diverse Social Images Task. Best result, human-based is in bold, and best second result, automatic run, is in italics.

run	P@20	CR@20	F1@20
run1	0.4180	0.3538	0.3637
run2	0.5367	0.4133	0.4425
run3	0.4305	0.3544	0.3745
run4	**0.5734**	**0.4252**	**0.4597**
run5	*0.5602*	*0.4179*	*0.4562*

Results are presented in table 2. It is interesting to point out that our best result for both precision and diversification is obtained with the multimodal human-based approach: estimating the probability of the relevance of each image to the query by a LLR model, and then ranking the final list with FCA clusters, *run4*, $F@20 = 0.4597$. Second best result, *run5*, $F@20 = 0.4562$, is the same *run4* approach, but automatically selecting the relevant and non-relevant images sets for estimating the LLR models. It is also important to observe that both automatic multimodal runs, *run5* and *run4*, overcome the automatic monomodal runs, *run1* and *run2* as expected.

4. CONCLUSIONS

We presented a multimodal approach, that estimates the relevance of the images by a relevance feedback algorithm using the visual features for determining the similarity. To handle image diversification, we apply a conceptual-based procedure (based on FCA and HAC) to cluster the images according to the latent topics addressed by their textual content. Results show that our multimodal approach works properly for retrieving similar diverse images. Results also show the importance of a proper selection of relevant and non-relevant sets for the relevance algorithm. A human knows better the meaning and the diversity of the topics. Our challenge is to make the approach automatic to be able to select these relevant and non-relevant images as a human being. The presented results are encouraging.

Acknowledgements. This work has been partially supported by PhD fellowship (FPI-UNED 2014), VOXPOPULI (TIN2013-47090-C3-1-P), MUSACCES (S2015/HUM3494) and DPI2013-47279-C2-1-R projects.

5. REFERENCES

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 5–14. ACM, 2009.

[2] F. Alqadah and R. Bhatnagar. Similarity measures in formal concept analysis. *Annals of Mathematics and Artificial Intelligence*, 61(3):245–256, 2011.

[3] X. Benavent, A. Garcia-Serrano, R. Granados, J. Benavent, and E. de Ves. Multimedia information retrieval based on late semantic fusion approaches: Experiments on a wikipedia image collection. *Multimedia, IEEE Transactions on*, PP(99):1–1, 2013.

[4] A. Castellanos, X. Benavent, A. García-Serrano, E. de Ves, and J. Cigarrán. Uned-uv @ retrieving diverse social images task. In *MediaEval Multimedia Benchmark Workshop, CEUR-WS.org, 1436, ISSN 1613-0073*, 2015.

[5] A. Castellanos, J. Cigarrán, and A. García-Serrano. Uned @ retrieving diverse social images task. In *MediaEval Multimedia Benchmark Workshop, CEUR-WS.org, 1263, ISSN 1613-0073*, 2014.

[6] E. de Ves, G. Ayala, X. Benavent, J. Domingo, and E. Dura. Modeling user preferences in content-based image retrieval: a novel attempt to bridge the semantic gap. *Neurocomputing*, 168:829–845, 2015.

[7] E. de Ves, X. Benavent, I. Coma, and G. Ayala. A novel dynamic multi-model relevance feedback procedure for content-based image retrieval. *Neurocomputing*, pages 99–107, October 2016.

[8] B. Ionescu, A. L. Gînscă, M. Zaharieva, B. Boteanu, M. Lupu, , and H. Müller. Retrieving Diverse Social Images at MediaEval 2016: Challenge, Dataset and Evaluation. In *MediaEval 2016 Workshop*, October 20-21, Hilversum, Netherlands, 2016.

[9] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[10] C. Loader. *Local regression and likelihood*. New York: Springer-Verlag, 1999.

[11] C. D. Manning, P. Raghavan, and H. Schütze. Hierarchical clustering. pages 377–403. 2008.

[12] S. Rudinac, A. Hanjalic, and M. Larson. Generating visual summaries of geographic areas using community-contributed images. *IEEE Transactions on Multimedia*, 15(4):921–932, 2013.

[13] R. Wille. Concept lattices and conceptual knowledge systems. *Computers & mathematics with applications*, 23(6):493–515, 1992.