

UNED-UV @ Retrieving Diverse Social Images Task

A.Castellanos
NLP & IR Group, UNED
C/ Juan del Rosal,16
Madrid, Spain
acastellanos@lsi.uned.es

X. Benavent
Department of Informatics,
University of Valencia
Valencia, Spain
xaro.benavent@uv.es

A. García-Serrano
NLP & IR Group, UNED
C/ Juan del Rosal,16
Madrid, Spain
agarcia@lsi.uned.es

E. de Ves
Department of Informatics,
University of Valencia
Valencia, Spain
esther.deves@uv.es

J.Cigarrán
NLP & IR Group, UNED
C/ Juan del Rosal,16
Madrid, Spain
juanci@lsi.uned.es

ABSTRACT

This paper details the participation of the UNED-UV group at the 2015 Retrieving Diverse Social Images Task. This year, our proposal is based on a multi-modal approach that firstly applies a textual algorithm based on Formal Concept Analysis (FCA) and Hierarchical Agglomerative Clustering (HAC) to detect the latent topics addressed by the images to diversify them according to these topics. Secondly, a Local Logistic Regression model, which uses the low level features and some relevant and non-relevant samples, is adjusted and estimates the relevance probability for all the images in the database.

1. INTRODUCTION

Information retrieval systems have been commonly based on maximizing the relevance of the result list (i.e., in terms of accuracy-based metrics). However, retrieval systems, and specially those focused on image diversification, should be able to offer relevant but also diverse results. Users are not only interested in accurate results but also in results covering different topics or situations [1].

To address this task we propose an image representation using the concept/s covered by the textual information related to the images. This conceptual representation is tackled by means of the use of Formal Concept Analysis, a data organization technique. In our participation in the 2014 edition of this task we proved that this approach was able to identify the different topics addressed in the images, allowing the diversification of the result list according to them [4].

This year we intend to go a step further by presenting a multimedia approach so that the aspects related to the visual information of the images were missed in our previous approach. As it has extensively been proven that the visual information has a great impact in the information retrieval systems [10]. The visual approach presented uses a relevance feedback algorithm developed by the UV group used in previous works [3], [5]. This method estimates the similarity probability of all the images of the database using the visual low-level features by means of a Local Logistic Regression

model [8].

2. SYSTEM DESCRIPTION

Our multi-media system has two sub-systems: a Textual-Based Information Retrieval system that works with textual information and generates clusters for diversity, and a Content-Based Information Retrieval sub-system that estimates the relevance of each of the images of the generated clusters.

2.1 Textual-Based Information Retrieval

Our proposal is based on the discovering of the latent topics addressed by the images by applying Formal Concept Analysis. A Hierarchical Agglomerative Clustering (HAC) [9] is then applied to group together similar images according to the detected *formal concepts* (those belonging to the same topic). Each HAC-based cluster may be considered as an image set covering a similar topic. Then, for each cluster, the visual features related to the images are applied to rank the cluster images according to their visual diversity.

2.1.1 FCA-based Modelling

Formal Concept Analysis (FCA) is a theory of concept formation [11] to organize *formal contexts*. A *formal context* is 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 the images in the test set are modelling 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) [7] 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.

2.1.2 HAC-based grouping

From the FCA *formal concepts*, a set of diverse image groups is created by applying a HAC algorithm [9]. Specif-

ically, 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 Content-Based Information Retrieval

This sub-system is concerned with Content-Based Image retrieval which models the user preferences by using a relevance feedback algorithm. The general methodology involves five steps:

1. Reduction of the data dimensionality: The provided low-level visual features [6] are used to generate a feature vector associated to each image that will be generically denoted as \mathbf{x} in a dimensional space $N = 945$. 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 $M < N$. 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 a few screens, each showing some images, and marks some of them as being relevant and non-relevant (*run5*). For the automatic runs (*run1* and *run3*), the relevant and non-relevant images are automatically selected. For the one-topic subset, the relevant images are the images given by Wikipedia for the certain topic, and for the multi-topic subset, we have generated the relevant images by selecting the first five results. A set of non-relevant images has been manually generated taking into account the non-relevant guidelines given [6] (photos with regular people as main subject, photos with riots and protests). At each query, the non-relevant images required are randomly selected from the generated non-relevant set.
3. Parameter estimation of the Local Logistic Regression Models [8]: The reduced feature vectors (PCA) and the relevant and non-relevant sets are the inputs of several Local Logistic Regression models whose outputs are the probabilities for user assessment, i.e. the probabilities he/she would assign to the fact that the 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 (\mathbf{p}) of dimension m for each individual image. We combine these probabilities in just one by using a weighted average. The weights (\mathbf{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 in the database.
5. Ranking of the database: the final diversity similarity rank is generated by selecting the highest probability

Table 1: Official Metrics for Retrieving Diverse Social Images Task. Best result for each topic set is in bold, and best result for the automatic runs is in italics. First block results are for the one-topic subset, the second for the multi-topic subset, and the third for the overall set.

Set		run1	run2	run3	run5
One-topic	P@20	0.6362	<i>0.7094</i>	0.7051	0.7645
	CR@20	0.3704	<i>0.4082</i>	0.3995	0.4194
	F1@20	0.4618	<i>0.5068</i>	0.4988	0.5240
Multi-topic	P@20	<i>0.7300</i>	0.6393	0.7207	0.7886
	CR@20	0.4257	<i>0.4407</i>	0.4116	0.4491
	F1@20	<i>0.5130</i>	0.5025	0.5001	0.5519
Overall	P@20	0.6835	0.6741	<i>0.7129</i>	0.7766
	CR@20	0.3983	<i>0.4246</i>	0.4056	0.4344
	F1@20	0.4876	<i>0.5046</i>	0.4994	0.5380

image at each of the clusters generated by the TBIR sub-system (*run3*). If there are less than 50 clusters, a second highest probability image selection is done. For the automatic run using only visual information (*run1*), the clusters are made by a k-means ($k = 50$) procedure over the PCA components of the visual feature vector.

3. RESULTS

We submitted four runs computed as following: *run1* - automated using visual information only (uses step 2 presented in Section 2.2), *run2* - automated using text information only (uses step 1 presented in Section 2.1), *run3* - automated multimedia (uses steps 1-2 presented in Sections 2.1 and 2.2) and *run5* - everything allowed: Textual clusters with FCA and manual relevance feedback algorithm using visual features (uses steps 1-2 presented in Sections 2.1 and 2.2). Results are presented in Table 1.

It is interesting to observe that our best results for both precision and diversification are obtained with the multimedia human-based approach, *run5*, $F@20 = 0.5380$, for both subsets: one-topic, $F@20 = 0.5240$, and multi-topic, $F@20 = 0.5519$. For the automatic runs, the best result is achieved by *run2* at the one-topic subset, $F@20 = 0.5068$; whereas for the multi-topic subset, *run1* gets the highest precision $P@20 = 0.7300$ and best performance $F1@20 = 0.5130$, but *run2* gets better diversification, $CR@20 = 0.4407$.

4. CONCLUSIONS

We presented a multimodal approach for image diversification applying a conceptual-based modelling (based on FCA and HAC) to cluster the images according to the latent topics addressed by their textual content, joined with a relevance feedback algorithm using the visual features for determining the similarity. Results show that the manual version of the multimedia approach works better than the automatic one. This is due to the way the relevant and non-relevant images are chosen to estimate the model. A human knows better the meaning of the topic; therefore, he/she selects the most significant images for the model. Our challenge is to make the automatic approach to be able to select the relevant and non-relevant images as a human being.

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, 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.
- [5] 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*, (0):–, 2015.
- [6] B. Ionescu, A. L. Gînsca, B. Boteanu, A. Popescu, M. Lupu, and H. Müller. Retrieving diverse social images at mediaeval 2015: Challenge, dataset and evaluation. In *Retrieving Diverse Social Images at MediaEval 2015: Challenge, Dataset and Evaluation. Working Notes Proceedings of the MediaEval 2015 Workshop*, 2015.
- [7] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [8] C. Loader. *Local regression and likelihood*. New York: Springer-Verlag, 1999.
- [9] C. D. Manning, P. Raghavan, and H. Schütze. Hierarchical clustering. pages 377–403. 2008.
- [10] 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.
- [11] R. Wille. Concept lattices and conceptual knowledge systems. *Computers & mathematics with applications*, 23(6):493–515, 1992.