REGIM@ 2016 Retrieving Diverse Social Images Task

Ghada Feki, Rim Fakhfakh, Noura Bouhlel, Anis Ben Ammar and Chokri Ben Amar

REsearch Groups on Intelligent Machines (REGIM), University of Sfax,
National Engineering School of Sfax (ENIS), Sfax, Tunisia
{ghada.feki, rim.fakhfakh, noura.bouhlel.tn, anis.benammar.tn, chokri.benamar}@ieee.org

ABSTRACT

In this paper, we describe our participation in the MediaEval 2016 Retrieving Diverse Social Images Task. The proposed approach refers to the Hierarchical, EM, Make Density-based clustering and the hypergraph-based learning to exploit visual, textual and user credibility-based descriptions in order to generate diversified results. We achieved promising results that our best run reached a F1@20 of 0.4105.

1. INTRODUCTION

Diversity is currently a hot research topic in the image retrieval context. The Retrieving Diverse Social Images task [1] deals with the problem of removing the redundant images from the top ranked images. Existing works always built clustering schemes to deal with this problem. The commonly used clustering techniques are hierarchical [2][3][4][5], spectral [6] and k-Medoids [7]. The general idea consists in returning images from different clusters in order to maximize the diversity rate among results. We propose a method considering both relevance and diversity. Indeed, based on the clustering techniques and the hypergraph-based learning, we exploit visual, textual and user credibility-based descriptions in order to generate diversified results.

2. PREVIOUS WORK

In 2012, our group within REGIM research laboratory participated in the Personal Photo Retrieval task [8] which also aims to diversify the results of the image retrieval systems by removing the redundant images from the top ranked images. The general idea of our proposed approach was based on browsing graphs which describe the similarity between images. Indeed, we generate an inter-images semantic similarity graph and an inter-images visual similarity graph. The retrieval process for each query takes into account not only the relevance-based ranking but also the diversity-based ranking which refers to the inter-images graphs to generate diversity scores. The approach was evaluated in more detail in [9] [10].

3. RUN DESCRIPTION

Among the participation in the Retrieving Diverse Social Images task, we focused on using the clustering techniques and the hypergraph-based learning since pairwise simple graphs used in our previous work scarcely represent relationships among images. Five runs were submitted as follows:

3.1 Run 1

The first run consists in the EM and Make Density-based Clustering [21] using only the visual information. The visual description which is used for this run is CNN generic. It is a descriptor based on the reference convolutional (CNN) neural network model which is learned with the 1,000 ImageNet[1] classes. The visual process contains four steps which are as follows. First, we estimate k number of clusters for each query by running EM clustering on the CNN generic description. Second, we carry out the make density-based clustering (wraps k-means algorithm) with a new empirical value of $k'=k+n$. Third, we extract the description of the cluster centers. Forth, based on the cosine similarity, we sort images by altering between the centers and choosing the closest one to selected center.

3.2 Run 2

The second run consists in the Hierarchical Clustering using only the textual information. Indeed, we have generated a hierarchical clustering over the corresponding image dataset based on textual descriptions [11][12]. The proposed hierarchical clustering contains four levels which are as follows:

- First level: Description
- Second level: Tags
- Third level: Title
- Fourth level: Location

3.3 Run 3

The third run consists in the combination of the EM and Make Density-based Clustering using the visual information (run 1) and the Hierarchical Clustering using the textual information (run 2) [13]. The final scores are generated as the mean average of the clustering-based visual scores and the textual scores [14].

3.4 Run 4

The fourth run consists of the combination between the aforementioned textual approach and a hypergraph-based visual approach. The visual description which is used for this run is CNN adapted. It is a descriptor based on the reference convolutional (CNN) neural network model which is learned with 1,000 tourist points of interest classes whose images were automatically collected from the Web. The hypergraph-based visual approach consists in the following steps. First, we use a hypergraph to model higher-order relationships between images. In such representation, the set of vertices denote the images for ranking. Each image is taken as a "*centroid*" vertex [15] and forms a hyperedge with its k-nearest neighbors. The Euclidean distance is used as a similarity function. Each hyperedege is weighted with a positive scalar denoting its importance in the hypergraph. Second, given the constructed hypergraph and an image query, we perform a hypergraph-based diverse ranking [16] algorithm with absorbing nodes to rank all remaining vertices in the hypergraph with respect to the query. The

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

[1] image-net.org

absorbing nodes [17] are used to avoid the redundant vertices from hitting higher ranking scores. The final scores are generated as the mean average of the hypergraph-based visual scores and the textual scores.

3.5 Run 5

In addition to the basic treatment provided by the run 3, the fifth run contains a refinement process which is based on the user credibility scores. Indeed, each image has a credibility-based score which consists in the mean average of the descriptors Visualscore, inverse of Faceproportion, tagSpecificity and meanImageTagClarity. The final ranking combines the user-based image ranking [18][19][20] with the ranking which is provided by the fusion of the EM and Make Density-based Clustering using the visual information and the Hierarchical Clustering using the textual information.

4. RESULTS AND DISCUSSION

This section presents the experimental results achieved on test set which contains 65 queries and about 19500 images. Table 1 shows the performance of the submitted runs according to both diversity and relevance. The used evaluation metrics are the Cluster Recall at X (CR@X) which is a measure that assesses how many different clusters from the ground truth are represented among the top X results, the Precision at X (P@X) which measures the number of relevant photos among the top X results and F1-measure at X (F1@X) which is the harmonic mean of the previous two.

Table 1. Run performance on Test Set

	Run1	Run2	Run3	Run4	Run5
P@20	0.5086	0.4797	0.5039	0.4852	0.5266
CR@20	0.36	0.3542	0.3501	0.3738	0.3702
F1@20	0.4024	0.3862	0.3964	0.4013	0.4105

Our clustering-based visual run (run 1) outperformed our textual run (run 2) in terms of all the metrics. The run which combined these two approaches performed better than the textual run in terms of P@20 but not in terms of CR@20. Indeed, comparing to the CR@20 achieved separately by the visual and the textual approaches, the combination decreases the diversity rate. However, with the refinement process which is based on the user credibility (run 5), all of the metrics had increased values. In fact, run 5 (clustering-based visual + textual + user credibility) outperformed run 1 (clustering-based visual), run 2 (textual), and run 3 (clustering-based visual + textual).

Similarly to the run 3, the run 4 is also a combined visual-textual run. Nevertheless, among the run 4, we have completely changed the visual approach. We notice that it outperformed the run 3 in terms of CR@20 but not in terms of P@20.

Thus, the best runs are the run 4 in term of CR@20 and the run 5 in terms of P@20 and F1@20. Consequently, we will detail more their results. As shown in figure 1, we notice that run 5 outperforms significantly the run 4 in term of Precision especially for the top ranked images. Concerning the Cluster Recall (Figure 2), we note that the two runs have close values. Until the top 10 ranked images run 5 outperforms slightly the run 4 and similarly by considering the top 50 ranked images run 4 outperforms slightly the run 5. Finally, with the F1-measure, we conclude that run 5 is almost the best run.

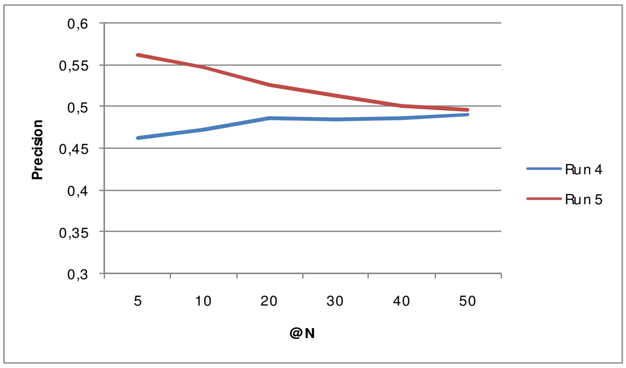

Figure 1. Precision comparison between Run 4 and Run 5

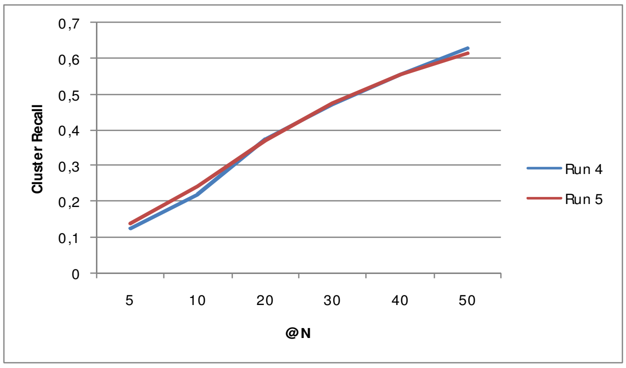

Figure 2. Cluster Recall comparison between Run4 and Run5

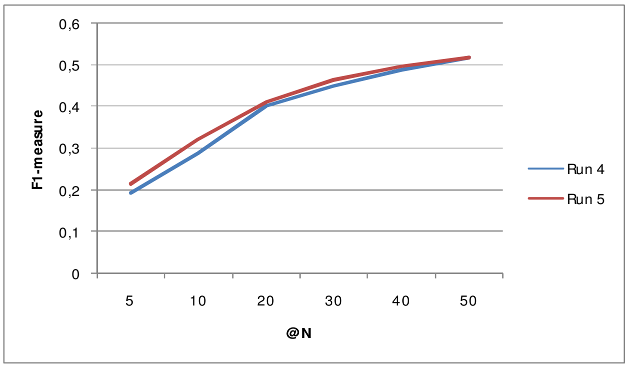

Figure 3. F1-measure comparison between Run 4 and Run 5

5. CONCLUSION AND FUTURE WORKS

Our participation in the MediaEval 2016 Retrieving Diverse Social Images Task achieved promising results. The proposed approach which refers to the clustering techniques and the hypergraph-based learning to exploit visual, textual and user credibility-based descriptions takes into account both relevance and diversity. Visual runs outperformed the textual one therefore further research will be mainly on the enhancement of the textual approach. Finally, we will also focus on the exploit of the user-credibility descriptors which seem to be very useful in the social media based retrieval.

ACKNOWLEDGEMENT

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

REFERENCES

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

[2] M. Zaharieva1, L. Diem, MIS @ Retrieving Diverse Social Images Task 2015. MediaEval 2015 Workshop, September 14-15, Wurzen, Germany, 2015.

[3] B. Boteanu, I. Mironica, B. Ionescu, LAPI @ 2015 Retrieving Diverse Social Images Task: A Pseudo-Relevance Feedback Diversification Perspective, MediaEval 2015 Workshop, September 14-15, Wurzen, Germany, 2015.

[4] X. Chen, H. Liu, Z. Deng, Y. Yang, Retrieving Relevant and Diverse Image from Social Media Images, MediaEval 2015 Workshop, September 14-15, Wurzen, Germany, 2015.

[5] A. Castellanos, X. Benavent, A. García-Serrano, E. de Ves, J. Cigarrán, UNED-UV @ Retrieving Diverse Social Images Task, MediaEval 2015 Workshop, September 14-15, Wurzen, Germany, 2015.

[6] Z. Paróczi, M. Kis-Király, B. Fodor, DCLab at MediaEval 2015 Retrieving Diverse Social Images Task, MediaEval 2015 Workshop, September 14-15, Wurzen, Germany, 2015.

[7] R. Calumby, I. Araujo, V. Santana, J. Munoz, O. Penatti, L. Li, J. Almeida, G. Chiachia, M. Gonçalves, R. Torres. Recod @ MediaEval 2015: Diverse Social Images Retrieval, MediaEval 2015 Workshop, September 14-15, Wurzen, Germany, 2015.

[8] D. Zellhofer, Overview of the Personal Photo Retrieval Pilot Task at ImageCLEF 2012, ImageCLEF 2012 Workshop.

[9] G. Feki, A. Ksibi, A. Ben Ammar, C. Ben Amar,"REGIMvid at ImageCLEF2012: Improving Diversity in Personal Photo Ranking Using Fuzzy Logic", In CLEF 2012.

[10] G. Feki, A. Ksibi, A. Ben Ammar, C. Ben Amar ,"Improving image search effectiveness by integrating contextual information, In 11th International Workshop on Content-Based Multimedia Indexing, CBMI 2013, pp149-154.

[11] G. Feki, A. Ben Ammar, C. Ben Amar "Adaptive Semantic Construction For Diversity-based Image Retrieval, Proceedings of the International Conference on Knowledge Discovery and Information Retrieval 2014,pp 444-449.

[12] A. Ksibi, G. Feki, , A. Ben Ammar, C. Ben Amar "Effective Diversification for Ambiguous Queries in Social Image Retrieval, 15th International Conference on Computer Analysis of Images and Patterns, CAIP 2013, pp571-578.

[13] R. Fakhfakh, G. Feki, A Ksibi, A. Ben Ammar, C. Ben Amar, “REGIMvid at ImageCLEF2012: Concept-based Query Refinementand Relevance-based Ranking Enhancement for Image Retrieval”, CLEF 2012

[14] G. Feki, R. Fakhfakh, A. Ben Ammar, C. Ben Amar, “Knowledge Structures: Which one to use for the query disambiguation?”, In 15th International Conference on Intelligent Systems Design and Applications 2015, pp 499-504.

[15] N. Bouhlel, A. Ksibi, A. Ben Ammar, and C. Ben Amar, “Semantic-aware framework for Mobile Image Search,” in 15th International Conference on Intelligent Systems Design and Applications (ISDA), 2015, pp. 479–484.

[16] Y. Huang, Q. Liu, S. Zhang, and D. Metaxas, “Image retrieval via probabilistic hypergraph ranking,” in Proc. IEEE Int. Conf. Comput.Vis. Pattern Recognit., Jun. 2010, pp. 3376–3383

[17] S. Sunderrajan and B. Manjunath, "Context-Aware Hypergraph Modeling for Re-identification and Summarization," IEEE Transactions on Multimedia, vol. 18, pp. 51-63, 2016

[18] R. Fakhfakh, A. Ksibi, A. Ben Ammar, C. Ben Amar, Fuzzy User Profile Modeling for Information Retrieval, KDIR 2014 (Knowledge Discovery and Information Retrieval), Rome, Italy, October 21-24, 2014: 431-436

[19] R. Fakhfakh, A. Ksibi, A. Ben Ammar, C. Ben Amar, Enhancing query interpretation by combining textual and visual analyses, International Conference on Advanced Logistics and Transport (ICALT), 2013, pp. 170-175

[20] G. Feki, R. Fakhfakh, A. Ben Ammar, C. Ben Amar. “Query Disambiguation: User-centric Approach”, In Journal of Information Assurance and Security, 2016 ,vol 11, pp. 144-156.

[21] T. S. Madhulatha, "An overview on clustering methods", IOSR Journal of Engineering, Apr. 2012, Vol. 2(4) pp:719-725.