MediaEval 2015: Retrieving Diverse Social Images with Image Search for Result Diversification

Shiran Dudy
Center for Spoken Language Understanding
OHSU
Portland, Oregon
dudy@ohsu.edu

Steven Bedrick
Center for Spoken Language Understanding
OHSU
Portland, Oregon
bedricks@ohsu.edu

ABSTRACT

These working notes will describe the motivation, process, results and analysis of results that the we have worked on as part of the MediaEval task of 'The 2015 Retrieving Diverse Social Images Task'. The concept of our approach was in implementing a technique [11] borrowed from documents retrieval field and applying it to the image domain with appropriate adjustments. The core idea here was that the decision making process, to produce the ranked image sequence, was done iteratively. Determining how different and relevant an image in the stack is relatively to the already chosen images.

Keywords

Information Retrival, Image Retrieval, Diversity function, Relevance, Ranked list

1. INTRODUCTION

Imagine you are in Munich and it's just about time that everybody around you talks about going to Oktoberfest. Being unfamiliar with this festival you are about to search for it to understand better if you'd like this event and what to expect. The task of this year's MediaEval 2015 was to provide the most diverse and relevant images to describe a place or an event in a spesific place given a query like Oktoberfest. The organizers provided us with a fully detailed task description alnog with data set for development and test found in [4].

2. RELATED WORK

The task of ordering images in a search engine given a query is still a developing field. The focus of this task is on retrieving diverse and relevant images from a given set of images. The motivation to our approach was based on a recent paper [11] that described an iterative scoring method for both relevance and diversity of a textual document. Every document was scored against the documents that were already chosen. The scoring function is described in Eq. 1:

$$f_s(x_i, R_i) = w_r^T \mathbf{x_i} + w_d^T h_s(R_i), \forall x_i \in X \backslash S \quad (1)$$

The scoring function combines information on relevance and diversity given the candidate document x_i and its diversity matrix R_i. While the prediction part above scores and chooses images the training part's purpose is to produce the relevance and diversity weight vectors w_r, w_d that are used in Eq. 1.

3. THE METHOD

Our task's objective is to utilize the scoring concept for images while incorporating the necessary tools to determine an image scoring function. In our task the relevence feature vector $\mathbf{x_i}$ was composed of Latent Semantic Analysis (LSA) [2] of 'tags' and 'description' textual fields, information on the user's credibility of 'visualScore', 1-'faceproportion', 'tagSpecificity', 'uniqueTags', 1-'locationSimilarity', and 1-'bulkProportion', and the normalized numerical data of image features. The diversity feature vector $h_s(R_i)$ was composed of the following features with their coresponding distance metrics: 'tags' and 'description' textual fields with cosinedisimilarity, this fields were used with Latent Dirichlet Allocation(LDA) [1] for topic diversity, 'csd' with l2 distance [8], 'hog' with Batacharia distance [9], 'cn' with euclidian distance [5], 'cm' with Canberra distance [3], 'lbp' with Chisquare distance [10], 'glr' with l1 distance [7], and 'cnn' with Mahalanobis distance [6]. A feature represent a more relevant or diverse as higher its value gets. The three algorithms described in [11] were reduced to two algorithms since queries were provided with ranked lists of relevance.

3.1 Settings

Settings that remained similar across experiments were: in relevance feature vector: user credibility information, lsa on 'tag' an 'description', and in diversity feature vector: lda on 'tag' an 'description', 'hog' and 'csd'. We implemented four different settings:
1) f1 - is using the global image features (without 'cnn')
2) f2 - is using the local image features (without 'cnn')
3) f1r f2d cnn - is using the global image features for relevance, local for diversity and global 'cnn'
4) f2 cnn2 - is using the local image features with local 'cnn'

4. RESULTS

Apart from reporting the official results we would first like to show the development set results of two experiments: f1

MediaEval '15 Wurzen, Germany
.

	F-score	RC	P
f1 Dev	0	0	0
f2 Dev	0.43	0.31	0.75

Table 1: Development set results.

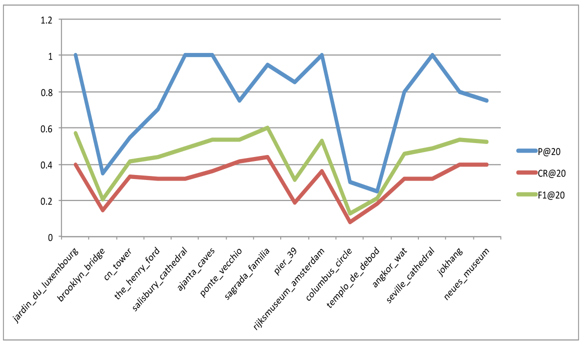

Figure 1: F-score, precision and coverage rates of f2 development.

and f2. These were computed seperately to the official results. We divided devset to 10-1 train test ratio and trained. Both f1 and f2 were trained and tested on same sets which were exclusice. Table 1 has the results of both experiments:

2) train, prediction, exp

graphs of the small set

5. CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LaTeX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

6. REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[2] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. *JAsIs*, 41(6):391–407, 1990.

[3] Z.-C. Huang, P. P. Chan, W. W. Ng, and D. S. Yeung. Content-based image retrieval using color moment and gabor texture feature. In *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, volume 2, pages 719–724. IEEE, 2010.

[4] B. Ionescu, A. Popescu, M. Lupu, A. L. Gınsca, and H. Müller. Retrieving diverse social images at mediaeval 2014: Challenge, dataset and evaluation. In *MediaEval 2014 Workshop, Barcelona, Spain*, 2014.

[5] H. Y. Lee, H. K. Lee, and Y. H. Ha. Spatial color descriptor for image retrieval and video segmentation. *Multimedia, IEEE Transactions on*, 5(3):358–367, 2003.

[6] S. Papadopoulos, A. L. Ginsca, A. Popescu, Y. Kompatsiaris, and I. Vlahavas. Improving diversity in image search via supervised relevance scoring.

[7] S. Selvarajah and S. Kodituwakku. Analysis and comparison of texture features for content based image retrieval. *International Journal of Latest Trends in Computing*, 2(1), 2011.

[8] T. Sikora. The mpeg-7 visual standard for content description-an overview. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6):696–702, 2001.

[9] R. YINZZ. Object tracking and detection after occlusion via numerical hybrid local and global model seeking. 1(8), 2008.

[10] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu. Boosting local binary pattern (lbp)-based face recognition. In *Advances in biometric person authentication*, pages 179–186. Springer, 2005.

[11] Y. Zhu, Y. Lan, J. Guo, X. Cheng, and S. Niu. Learning for search result diversification. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 293–302. ACM, 2014.