

Re-ranking for Multimedia Indexing and Retrieval

Bahjat Safadi, Georges Quénot

▶ **To cite this version:**

Bahjat Safadi, Georges Quénot. Re-ranking for Multimedia Indexing and Retrieval. ECIR 2011 - European Conference on IR Research, Apr 2011, Dublin, Ireland. pp.708-711, 10.1007/978-3-642-20161-5_76 . hal-00763562

HAL Id: hal-00763562
https://hal.science/hal-00763562

Submitted on 11 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Re-ranking for Multimedia Indexing and Retrieval

Bahjat Safadi and Georges Quénot

Laboratoire d'Informatique de Grenoble
BP 53 - 38041 Grenoble Cedex 9, France
{Bahjat.Safadi, Georges.Quenot}@imag.fr

Abstract. We proposed a re-ranking method for improving the performance of semantic video indexing and retrieval. Experimental results show that the proposed re-ranking method is effective and it improves the system performance on average by about 16-22% on TRECVID 2010 semantic indexing task.

Keywords: Multimedia Indexing and Retrieval, Re-ranking.

1 Introduction

Semantic indexing and retrieval for multimedia databases has been a very active research field over the past few years. The global goal of multimedia indexing is to automatically describe documents containing images, sounds or videos. Nevertheless, allowing users to retrieve them within large collections or to easily navigate in these collections is still a very hard task.

Semantic indexing is generally achieved by supervised learning where a system is trained on positive and negative samples of a target concept (the development set) for producing a model which is then used for predicting the likeliness of new samples to contain this concept (the test set). This likeliness is often computed homogeneously to a probability for each data sample to contain the concept. Retrieval can then be done by ranking the samples according to the probability score. Such ranking is initially done with a score independently for each sample using only information from the development set. It is often possible to improve the indexing or retrieval performance by re-scoring the samples using the initial scoring information on the whole test collection. Recently, several ways of re-ranking methods have been proposed and developed, below we reviewed some of these methods.

Context fusion [1, 2]: the results of different searching models (concept-based search model, text-based search model and query by example) are used to re-rank the ranked lists, in fact here the focus is on the fusion of different model outputs. This method needs to train new classifiers on new descriptors. Since we also use in our work the fusion of the outputs from multiple models, we took this as a baseline approach.

Classification-based re-ranking [3]: the initial results of a baseline system are used to discover the co-occurrence patterns between the target semantics and extracted features. This is very similar to *"learning to rank"*[4], which is based on training a ranking model which can precisely predict the ranking lists in the dataset. In [3] the authors used the top-ranked and bottom-ranked samples respectively, as pseudo-positive and pseudo-negative examples to train a new classification model for ranking, and the classification

margin for a target concept is regarded as its (new) re-ranked. The use of SVM as the classification model, leads to the method called *RankSVM*[4].

Ordinal re-ranking, in [6]: the author re-ranks an initial results by using the co-occurrence patterns via the ranking functions. The final score is the weighting combination of the original score and the re-ranked scores. They adopted a training method to train the Re-ranking algorithm on some concepts, and the re-ranking algorithm was applied to re-rank the remaining concepts.

In the case of video collections, the retrieval units are often not the whole videos themselves (which are generally too coarse grain for the user needs) but the video shots composing the videos. Our contribution in this paper, is to re-rank the video shots according to their scores, which were obtained from the classifiers, according to the video knowledge and nature. Our work is similar to the work in [5], where the authors re-ranked the previous results with video knowledge which is described as the mean score value of the current shots in the same video. The mean scores were adopted to be fused with the original scores.

The paper is organized as follows: Our re-ranking method is presented in section 2. Section 3 describes the experimental results, while section 4 presents our concluding remarks.

2 Re-ranking method

In this paper, we propose an unsupervised method for re-ranking video shots according to a query or a concept. Our hypothesis is that videos have rather homogeneous contents and that the presence of a given concept in a video depends a lot on the nature of the video itself. Scores (here homogeneous to probabilities) are computed independently for all video shots as their likeliness to contain a target concept using classifiers (or networks of classifiers) that were trained on the development set. The re-ranking is actually done by a re-scoring which is done in two steps. First, we compute a global score for each video for containing the target concept; this score is computed from the scores of all the shots within the video. Then, we re-evaluate the score of each shot according to the global score of the video it belongs to.

The test collection contains a set of videos $V = (v_1, v_2, \ldots, v_m)$, m being the number of videos in the collection. Each video v_i composed of a sequence of shots $v_i = (s_{i1}, s_{i2}, \ldots, s_{in_i})$, n_i being the number of shots of v_i.

For each shot s_{ij}, an initial classification score x_{ij} is computed from supervised learning on the development set. Many options are possible for the computation of a global score x_i for a video v_i from the shots that it contains. We tried several formulas and found that the following one which is a generalization of the mean of the shot scores was the most effective:

$$z_i = \left(\frac{\sum_{j=1}^{n_i} (x_{ij})^\alpha}{n_i} \right)^{1/\alpha}, \tag{1}$$

where α is the parameter that has to be tuned by cross-validation within the development collection. Then, we update the score of each shot according to its previous score and

the global score of the video it belongs to. Again, many options were possible and we chose a weighted multiplicative fusion:

$$x'_{ij} = x_{ij}^{1-\gamma} \times z_i^{\gamma}, \tag{2}$$

where γ is a parameter that controls the "strength" of the re-ranking. It also has to be tuned by cross-validation within the development collection.

3 Experiments

We conducted our experiments on TRECVID 2010, where 130 concepts are provided with ground truth labels in a training set. The evaluation is done by calculating the Mean Average Precision (MAP) on only 30 concepts that were chosen by NIST. We evaluated the re-ranking method on four different initial classification results, which we have submitted to TRECVID 2010, including different fusion strategies such as weighted and direct optimized weighted fusion, also the combination of the two fusion types with genetic fusion. These fusion strategies were applied on score vectors obtained by training different systems on 45 different descriptors (audio and visual descriptors) which have been produced by various partners of the IRIM project of the GDR ISIS [7].

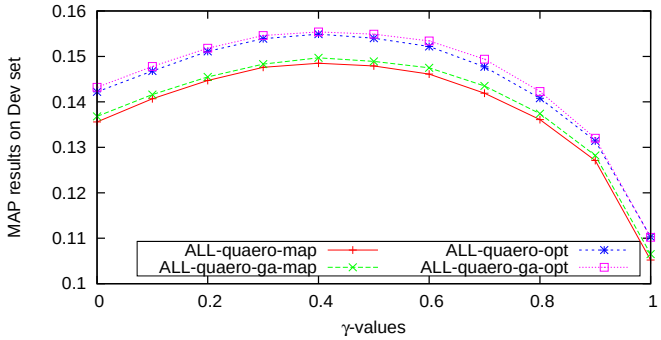


Fig. 1. Tuning of the γ parameter: re-ranking the results after fusion with four different runs; MAP on the 130 concepts on the development set of TRECVID 2010.

As it is shown, in order to get the best performance from our system, we need to tune the γ parameter in our re-ranking algorithm; this tuning was conducted on the development set using four different runs that had different fusion strategies. Fig. 1 shows the results of our re-ranking method on the development set using different values for γ. Each one of the curves indicates the MAP on the validation set after re-ranking the fused results with different values of γ. Here the baseline, for each run (curve), is given when $\gamma = 0$. As we can see, for the four runs, the best performance is reached when $\gamma = 0.4$; here the re-ranking is applied directly on the fused results.

We made initial experiments with $\alpha = 1$ (regular mean) and then tried other α values. We found a small improvement with higher values with an optimal value close to 2, corresponding to a root mean square. We then applied the proposed method on

the TRECVID 2010 test set with $\gamma = 0.4$ and $\alpha = 2$. Table. 1 shows the gain on the MAP using our re-ranking method on four different initial scoring methods. As we can see, our proposed re-ranking method can significantly improve the performance: on this collection the gain is up to 22%.

Run	Baseline	Re-rank	Gain(%)
ALL_quaero_map	0.0476	0.0577	21
ALL_quaero_ga_map	0.0479	0.0584	22
ALL_quaero_opt	0.0485	0.0563	16
ALL_quaero_ga_opt	0.0484	0.0568	17

Table 1. Results obtained on the test set; MAP on the 30 concepts of our four submitted results for TRECVID 2010.

4 Conclusion

We proposed a re-ranking method which improves the performance of semantic video indexing and retrieval. Experimental results show that the proposed re-ranking method is effective and that it can improve the performance of our system on average by about 16-22% on TRECVID 2010 semantic indexing task.

Acknowledgments

This work was partly realized as part of the Quaero Program funded by OSEO, French State agency for innovation.

References

1. W. Jiang et al, Context-based concept fusion with boosted conditional random fields, IEEE ICASSP, 2007.
2. J. Liu, W. Lai, X.S. Hua, Y. Huang, and S. Li, Video Search Re-Ranking via Multi-Graph Propagation, Proc. 15th Int'l Conf. Multimedia, pp. 208-217, 2007.
3. L. Kennedy and S.-F. Chang, A reranking approach for context-based concept fusion in video indexing and retrieval, ACM CIVR, pp. 333-340, 2007.
4. R. Herbrich, T. Graepel, and K. Obermayer, Support vector learning for ordinal regression, ICANN, pp. 97-102, 1999.
5. F. Wang and B. Merialdo, Eurecom at TRECVID 2009 High-Level Feature Extraction, TRECVID 2009 Workshop Agenda, Gaithersburg, MD USA, Nov 2009.
6. Yi-Hsuan Yang; Hsu, W.H.; , Video search reranking via online ordinal reranking, Multimedia and Expo, 2008 IEEE International Conference on , pp.285-288, 2008
7. David Gorisse, Frédéric Precioso, Philippe Gosselin, Lionel Granjon, Denis Pellerin, Michèle Rombaut, Hervé Bredin, Lionel Koenig, Hélhe Lachambre, Elie El Khoury, Rémi Vieux, Boris Mansencal, Yifan Zhou, Jenny Benois-Pineau, Hervé Jégou, Stéphane Ayache, Bahjat Safadi, Yubing Tong, Franck Thollard, Georges Quénot, Alexandre Benoît, Patrick Lambert, IRIM at TRECVID 2010: High Level Feature Extraction and Instance Search. In *TREC Video Retrieval Evaluation workshop*, Gaithersburg, MD USA, November 2010.