Image Re-ranking based on Relevance Feedback Combining Internal and External Similarities

R. Omar Chávez García, Manuel Montes and L. Enrique Sucar
Instituto Nacional de Astrofísica Óptica y Electrónica
Luis Enrique Erro Num. 1, Sta. María Tonantzintla, Puebla, México
{romarcg, mmontes, esucar}@ccc.inaoep.mx

Abstract

We propose a novel method to re-order the list of images returned by an image retrieval system (IRS). The method combines the original order obtained by an IRS, the similarity between images obtained with textual features and a relevance feedback approach, all of them with the purpose of separating relevant from irrelevant images, and thus, obtaining a more appropriate order. Experiments were conducted with resources from Image CLEF 2008; the proposed method improves the order of the original list up to 42%.

Introduction

An image retrieval system (IRS) receives a query from a user, as keywords or sample images, and it is expected to return an ordered list of images that satisfies the user's request. Ideally the IRS should return a list with all relevant images ordered according to the user's request. Current IRS, in general, tend to include several relevant images in the retrieved list. However, the images are not ordered properly. That is, IRS have a relatively good performance in terms of *recall*, but poor in terms of *precision*.

One way to improve the order of the results of an IRS is to use *relevance feedback*. Some approaches for relevance feedback attempt to enrich the query to perform a new retrieval and obtain better results, but this can be computationally expensive. This motivates to use only the retrieved list and reorder the images on the assumption that this list has relevant images, but not necessarily in the first positions.

Previous work (Cui, Wen, and Tang 2008; Deselaers et al. 2008) do not use all the information available. We consider that all the available information –the original order, the subset obtained via relevance feedback, the original query, and the entire list of retrieved images– is useful to improve the list order, and we propose a re-ranking method that combines all this information to obtain a more appropriate order, based on a Markov random field (MRF) model.

Experiments were conducted using the resources of ImageCLEF 2008. We used one of the IRS that participated in this forum as the base retriever to obtain the initial lists. Each of the results obtained by our method improved the original list order; the improvements obtained are up to 42%.

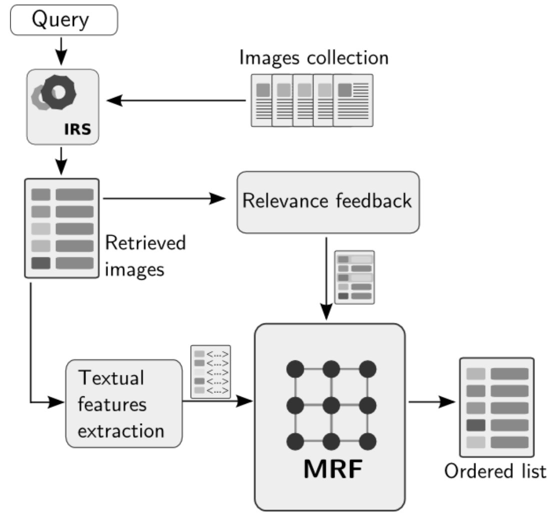

Figure 1: Block diagram of the proposed method.

Proposed Method

A general outline of the proposed method is given in Figure 1. Given a query, the IRS retrieves from a given collection of images (that includes text captions) a list of images sorted according to a relevance criteria. From this list, some relevant images are selected based on a relevance feedback approach. The textual description of each image, the query given by the user, and a subset of images selected via relevance feedback, are combined to produce a re-ordered list. This re-ranking is obtained based on a MRF model that separates the relevant images from irrelevant ones.

We consider a MRF in which each node corresponds to an image (text caption) in the list. Each image is represented as a random variable with 2 possible values: *relevant* and *irrelevant*. We consider a fully connected graph, such that each node (document) is connected to all other nodes in the field. The energy function of the MRF combines two factors: the similarity between the images in the list (*internal* similarity); and information obtained from the original order and the similarity of each image with the query (*external* similarity). The proposed energy function takes into account both aspects and is defined as follows: $V_c(f) + \lambda V_a(f)$. Where

V_c is the interaction potential and it considers the similarity between f and its neighbors. V_a is the observation potential and represents the influence of external information. The weight factor λ favors V_c ($\lambda < 1$) or V_a ($\lambda > 1$). V_c is defined as:

$$V_c(f) = \begin{cases} \bar{Y} + (1 - \bar{X}) & if\ f = irrelevant \\ \bar{X} + (1 - \bar{Y}) & if\ f = relevant \end{cases}$$

Where $\bar{Y}$ represents the average distance between f and its irrelevant neighbors. $\bar{X}$ represents the average distance between f and its relevant neighbors. The distance metric used to measure similarity is defined as: $1 - sim(f, g)$, where $sim(f, g)$ is defined as: $sim(f, g) = \frac{2|f \cap g|}{|f \cup g|}$. V_a is defined as follows:

$$V_a(f) = \begin{cases} (1 - dist(f, q)) \times h(posinv(f)) & iff = irrelevant \\ dist(f, q) \times h(pos(f)) & iff = relevant \end{cases}$$

The V_a potential is obtained by combing two factors. The first indicates how similar, $dist(f, q)$, or different, $1 - dist(f, q)$ is f from the query q. Where $dist(f, q)$ is defined as: $dist(f, q) = |f \cap q|/|q|$. The second is a function that converts the position in the list given by a base IRS to a real value. The function $pos(f)$ and $posinv(f)$ returns the position and inverse position of the image f in the original list respectively.

The initial configuration of the MRF is obtained by relevance feedback: the images selected by the user are initialized as relevant, and all other as irrelevant. Then, the MRF configuration of minimum energy (MAP) is obtained via stochastic simulation using the ICM algorithm. At the end of this optimization process, each variable (image) has a value of relevant or irrelevant. Based on these values, a new re-ordered list is produced, by positioning first the relevant images and then the irrelevant ones.

Experimental Results

We conducted a series of experiments to evaluate the proposed method, compared with the original list, and its sensitivity to the model parameters. We used the resources of Image CLEF 2008, which consist of the image collection IAPR TC-12, a set of 39 queries for the photo retrieval track and a list of results from one of the participants (TIA-TXTIMG). These queries have 3 sample images and a textual description that includes a narrative about the images relevant to the query. The TIA-TXTIMG SRI retrieves a list of images by combining the results of several retrieval methods (Escalante et al. 2008). We considered the best results obtained by this group as the input for our method, selecting the first 100 images retrieved by TIA-TXTIMG IRS for each of the 39 queries.

For representing the images we used a binary bag of words representation, in which each vector element represents a word from the collection vocabulary; the query is represented in the same manner. Each of the images in the collection has assigned a set of descriptive fields, we included the words in the title and in the textual description to represent the images. For evaluation we used the MAP.

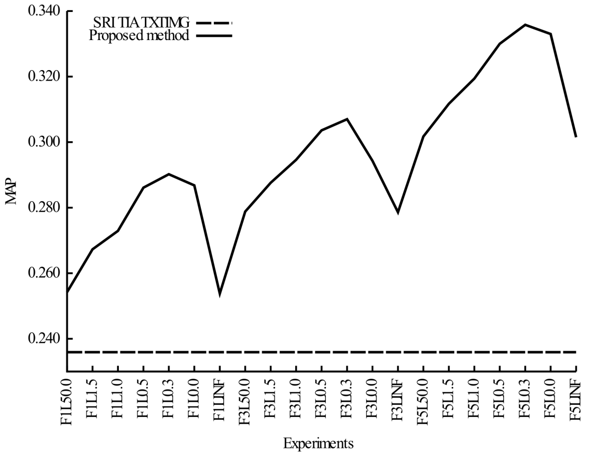

Figure 2: MAP obtained by the proposed method (Y axis) for different configurations (X axis): F– number of images taken as feedback, L–value of λ (the bottom line shows the MAP of the original list).

Five experiments were conducted varying λ, and considering 1, 3 and 5 images as relevance feedback. We used a *simulated* user feedback as we know beforehand the relevant images. Figure 2 shows a comparison between the results from the original list and the results obtained with the proposed method for different configurations of parameters. The results show that considering only textual attributes to measure the similarity, an improvement of up to 42% is obtained selecting 5 images as feedback, and an improvement of 23% when only 1 image is selected as feedback; as more images are given as feedback, the performance improves. When the value of λ is small (eg. 0.3) the proposed method yields the best results.

Conclusions and Future Work

Experiments in the ImageCLEF 2008 Photo retrieval task show that our method obtains a significant improvement with respect to the original list; the improvement increases proportionally to the feedback quantity, and the best results are obtained by giving a higher weight to the internal similarity. As future work we plan to combine textual and visual features to exploit the advantages of both.

References

Cui, J.; Wen, F.; and Tang, X. 2008. Real time google and live image search re-ranking. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, 729–732. New York, NY, USA: ACM.

Deselaers, T.; Paredes, R.; Vidal, E.; and Ney, H. 2008. Learning weighted distances for relevance feedback in image retrieval. In *International Conference on Pattern Recognition 2008*.

Escalante, H. J.; Gonzalez, J. A.; Hernandez, C.; Lopez, A.; Montes, M.; Morales, E.; Sucar, E.; and Villasenor, L. 2008. Tia-inaoes participation at imageclef2008. In *Working Notes of the CLEF 2008 Workshop*.