Image Retrieval with Relevance Feedback:
From Heuristic Weight Adjustment to Optimal Learning Methods

Thomas S. Huang,   Xiang Sean Zhou
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana Champaign, Urbana, IL 61801, USA
{huang, xzhou2}@ifp.uiuc.edu

Abstract

Various relevance feedback algorithms have been proposed in recent years in the area of content-based image retrieval. This paper gives a brief review and analysis on existing techniques—from early heuristic-based feature weighting schemes to recently proposed optimal learning algorithms. In addition, the kernel-based biased discriminant analysis (KBDA) is proposed to fit the unique nature of relevance feedback as a biased classification problem. As a novel variant of traditional discriminant analysis, the proposed algorithm provides a trade-off between discriminant transform and regression. The kernel form is derived to deal with non-linearity in an elegant way. Experimental results indicate that significant improvement in retrieval performance is achieved by the new scheme.

1. Introduction

In spite of the significant advances in machine intelligence and computer vision technology, fully automated image understanding by machines is still an unsolved problem. In the context of content-based image retrieval (CBIR), even if we assume that consensus interpretation of images can be reached among all possible users at all times (which is seldom true), the gap exists between the high level semantics in the human minds (whether expressible by words or not) and the low level features computable by machines. So the first challenge for the machine is to *learn the associations between high-level and low-level* to effectively facilitate retrieval. However, "A picture is worth a thousand words"—a more profound challenge comes from the dynamic interpretation of images under various circumstances. In other words, the perceptual "similarity" depends upon the application, the person, and the context of usage. So in fact the machine not only needs to learn the associations, but also has to learn them *on-line with user in the loop*. A natural way of getting user in the loop is to ask the user to give feedbacks regarding the relevance of the current outputs of the system. Though this is an idea borrowed from the text retrieval field [14], it seems to work better in image domain: it is easier to tell the relevance of an image than that of a document—it takes time to read through a document while an image reveals its content instantly.

Aside from the *real-time* demand, another difficulty for learning during user feedback is the relatively small number of training samples (comparing to traditional requirements) during the user interaction, usually only 10 to 20 per round for an image retrieval system. Often times this is even smaller than the dimension of the feature space. These difficulties make learning in CBIR a unique, challenging, yet exciting research area.

In this paper, we use the term "relevance feedback" to denote the on-line learning process during information retrieval, based on the relevance judgments fed-back by the user. Various relevance feedback techniques have been shown to provide dramatic performance boost in CBIR systems [10][12][4][17][19].

2. State of the Art: Review and Analysis

2.1 From Heuristic to Optimal Scheme

In its short history, relevance feedback developed along the path from heuristic based techniques to optimal learning algorithms, with early work inspired by term-weighting and relevance feedback techniques in document retrieval [14]. These methods proposed heuristic formulation with empirical parameter adjustment, mainly along the line of independent axis weighting in the feature space [10][12][9][11][15]. The intuition is to emphasize more on the feature(s) that can best cluster the positive examples and separate the positive and the negative.

Early works [10][12] have clear birthmarks from document retrieval field. For example, Picard et al. [10] quantizes the features and then groups the images or regions into hierarchical trees whose nodes are constructed through single-link clustering. Then weighting on groupings is based on "set operations". In [12], learning based on "term frequency" and "inverse document frequency" in text domain is transformed into learning based

on the ranks of the positive and negative images along each feature axis in the continuous feature space.

An interesting scheme using SOM (Self-organizing Map) is worth noting[5]: it uses TS (Tree-Structured)-SOMs to index the images along different features. Positive and negative examples are mapped to positive and negative impulses on the maps and a low-pass operation on the maps is argued to implicitly reveal the relative importance of different features because a "good" map will keep positive examples cluster while negative examples scatter away. This is based on similar intuition as that of [9], where a probabilistic method is used to capture feature relevance.

Aside from their lack of optimality claim, the assumption of feature independence imposed in these heuristic methods is also forceful, unless independent components can be effectively extracted beforehand.

Later on researchers begin to look at this problem from a more systematic point of view by formulating it into an optimization, learning, classification, or density estimation problem. In [4] and [13], based on the minimization of total distances of positive examples from the new query, the optimal solutions turn out to be the weighted average as the new query and a whitening transform (or Mahalanobis distance). (The pseudo-inverse treatment of a singular covariance matrix in [4], Appendix D will lead to contradictory and counter-intuitive results for zero-variance subspace.) Rui and Huang [13] adopt a two-level weighting scheme (with partial independence assumption) to better cope with singularity issue caused by the small number of training samples. To take into account the negative examples, Schettini et al. [16] updates the feature weights along each feature axis by comparing the variance of positive examples to the variance of the union of positive and negative examples. This is an interesting intuition yet falls short of formal justification.

Assuming that the user is searching for a particular target, and the feedback is in the form of "relative judgment", the PicHunter [2] proposes the stochastic comparison search as its relevance feedback algorithm.

MacArthur et al. [6] cast relevance feedback as a two class learning problem, and used a decision tree algorithm to sequentially "cut" the feature space until all points within a partition are of the same class. The database is classified by the resulting decision tree: images that fall into a relevant leaf are collected and the nearest neighbors of the query are returned.

While most CBIR systems use well-established image features such as color histogram/moments, texture, shape, and structure features, there are alternatives. Tieu and Viola [17] use more than 45,000 "highly selective features", and a boosting technique to learn a classification function in this feature space. In [18], Gaussian mixture model on DCT coefficients is used as image representation. Then Bayesian inference is applied for image regional matching and learning.

Recently there are also attempts to incorporate support vector machine (SVM) into relevance feedback process [1][3]. However, SVM as a two-class classifier is not directly suitable for relevance feedback, because the training examples are far too few to be representatives of the true distributions. However a kernel based one-class SVM as density estimator for positive examples has been shown to outperform the whitening transform based linear method [1].

Without assuming one Gaussion mode for positive examples, Parzen window density estimation can be applied to capture non-linear, multi-mode distribution of positive examples [7], and an "aggregate dissimilarity" function is used to combine for a candidate test image the pairwise distances to every positive examples. The major weakness of these density estimation based schemes is in their lack of discriminative power to exclude negative examples with good generalization capabilities.

Formulated in the transductive learning framework, D-EM algorithm [20] uses examples from the user feedback (labeled data) as well as other data points (unlabeled data). It performs discriminant analysis inside the EM iterations to select a good subset of features and to relax the assumption of probabilistic structure of data distributions. Noting that performing discriminant analysis alone without EM will fail miserably due to its strong assumption on the negative examples, and a better alternative is proposed in [21], for which we will give some detailed discussion in the following sections and propose its kernel-based version to take into account non-linearity in the original feature space.

2.2 Variants and Assumptions

It should be noted that not all the algorithms discussed above are directly comparable: more often than not, different schemes have different assumptions, such as

a. "What is the user looking for?" Some assumes the user is looking for "a particular target image" [2], while many others assume the user is looking for *"similar" images* to the query (or sketch) at hand [3][4][12][13].
b. "What to feedback?" Some algorithm assumes the user will give a binary feedback for positive and negative examples [17]; some only takes positive examples [4]; some takes *positive and negative examples with "degree of (ir)relevance"* for each [12]; some assumes the feedback to be a comparative judgment, i.e., the positive examples are "more like the target than the negative ones" [2].
c. "What's the distribution?" Another issue is what assumption to be imposed on the target image class(es). Gaussian assumption is the most common and convenient one [4]. However, recent *kernel based algorithms* can deal with non-linearity in an elegant way [1]. (Also see section 4.)

d. "What to learn, and how?" A majority of the work uses relevance feedback to learn a *new query* and the relative importance of different features[9][12][15], with some tries to learn a linear *transformation in the feature space* either with [4][13] or without [12] considering correlations among feature components. While others treat it either as a density estimation [1][7], learning [18][20][19], or classification [17][21] problem.

For each of the above questions, the *phrases in italic* represent our answers for a new relevance feedback scheme proposed in this paper (see Section 3 and 4). In other words, we assume that the user is looking for a class of images, with feedback in the form of a real number from the interval [-1,1]; and we try to model non-linear distributions; and the system learns a new query and a non-linear discriminating transform in the feature space before nearest neighbor retrieval is performed in the new space.

2.3 Relevance Feedback as a Biased Classification Problem

Much work regards the relevance feedback problem as a strict two-class classification problem, with equal treatments on both positive and negative examples. It is reasonable to assume positive examples to cluster in certain way (maybe non-linearly), but negative examples usually do not cluster since they can belong to any class. Forcefully assigning all negative examples into one class/mode can mislead the algorithm therefore hurt the robustness in performance, especially when the number of training samples is small.

We define *biased classification problem* as the problem in which there are an unknown number of classes but the user is only interested in one class, i.e., the user is *biased toward one class*. And the training examples are labeled by the user as only positive or negative as to whether they belong to the target class or not. Thus the negative examples can come from an uncertain number of classes. Much research simply treat this case as a two-class classification problem, i.e., the negative examples are all treated as one class. However the intuition is that "the positive examples are all good in the same way, but bad examples are bad in their own ways". More importantly the limited number of negative examples are poor representations of the true distribution of ALL negative points. (One alternative is to use a random sampling strategy to increase the number of negative examples thus their representative power [17]. But this carries the danger of mislabeling.) Therefore it is desirable to distinguish a *real* two-class problem from a biased classification (or *(1+x)-class*) problem in our case. And a typical relevance feedback process poses a biased classification problem.

For a biased classification problem, we ask the following question: what is the optimal linear trans-formation to take into account both positive and negative examples, such that positive examples have minimal scatter while negative examples have maximal scatter with respect to positive ones (i.e., negative examples are "far away" from positive ones)? Or mathematically, what is the optimal transformation such that the ratio of "positive scatter" over "negative scatter with respect to positive" is minimized? We call this biased discriminant analysis (BDA) due to the bias toward the positive examples [21].

3. Biased Discriminant Analysis (BDA)

We use $\{x_i\}$, $i=1,...,N_P$ to denote the positive examples, and $\{y_i\}$, $i=1,...,N_N$ to denote the negative examples. And m_x is the mean vector of $\{x_i\}$. For any linear transformation matrix W, which is a long matrix, we define the biased criterion function

$$J(W) = \frac{|W^T S_N W|}{|W^T S_P W|} \quad (1)$$

where

$$S_N = \sum_{i=1}^{N_N} (y_i - m_x)(y_i - m_x)^T \quad (2)$$

and

$$S_P = \sum_{i=1}^{N_P} (x_i - m_x)(x_i - m_x)^T \quad (3)$$

The optimal biased linear transformation is

$$W^o = \arg\max_W \{J(W)\} \quad (4)$$

To avoid singularity in estimating the covariance matrix in the solution, regularization shall be applied by adding small positive values to the diagonal before inversion. The influence of the negative examples can be tuned by adding a discounting factor—at the extreme value of this factor this scheme is reduced to the whitening transform on positive examples.

4. Kernel-based Nonlinear BDA (KBDA)

To take into account non-linearity in the data, we propose a kernel-based approach (*cf.* [8]). The original BDA problem is reformulated into dot-product form, and then a kernel is used in place of the dot-product. This is the major advantage of kernel-based approach—if we were to carry out the nonlinear transformation explicitly before the dot-product calculation, the computation would usually be formidable. Assume a nonlinear transform Φ on the original data before applying linear BDA:

$$J(W) = \frac{w^T S_y^\phi w}{w^T S_x^\phi w} \quad (5)$$

It can be shown that the numerator of (5) can be re-written as:

$$w^T S_y^\phi w = \alpha^T (K_y - K_x I_{N_x}^y)(K_y - K_x I_{N_x}^y)^T \alpha \quad (6)$$

where:

$$K_{y_j} = \Phi^T \phi(y_j), \quad K_{mx} = \Phi^T m_x^\phi, \quad (K_y)_{:,j} = K_{y_j} \quad (7)$$

and $I_{N_x}^{y}$ is an N_x by N_y matrix of all elements being $1/N_x$.

Similarly, rewrite the denominator of (5),

$$w^T S_x^{\phi} w = \alpha^T K_x (I - I_{N_x}^{x}) K_x^T \alpha \qquad (8)$$

and $I_{N_x}^{x}$ is an N_x by N_x matrix of all elements being $1/N_x$.

With optimal $\boldsymbol{\alpha}$, the projection of a new pattern z onto w is given by:

$$w^T \phi(z) = \sum_{i=1}^{N_x} \alpha_i k(x_i, z) + \sum_{j=1}^{N_y} \alpha_{j+N_x} k(y_j, z) \qquad (9)$$

5. Experimental Results

Extensive experiments have demonstrated the advantages of the proposed algorithm. Here we can only report one experiment on a fully labeled set of 500 images from COREL. It contains five classes, each with 100 images. Features used are color moments, wavelet moments, and water-filling structural features [22]. Each round 10 positive and 10 negative images are randomly drawn as

Figure 1. Averaged Error Rates

training samples and the learned feature transformation is applied to all the images before a nearest neighbor classifier is applied for image ranking. For each round the error rate in the top 100 returns is recorded as the performance measures. A total of 500 rounds of testing are performed and the averaged error rates are shown in Figure 1, where four schemes are compared: WT (whitening transform on positive examples), DA (Fisher discriminant analysis), BDA, and KBDA. One can see KBDA outperforms others on average by a significant margin.
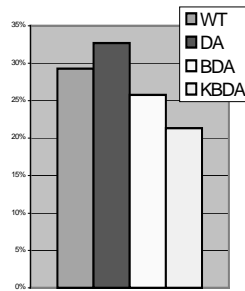
6. Conclusion

In this paper, we have briefly reviewed the existing relevance feedback techniques. Emphasize is put on the analysis of the unique characteristics of relevance feedback as a learning/classification problem. A new scheme is also proposed with experimental results supporting its superior performance than existing schemes.

Reference

[1]   Y. Chen, X. S. Zhou, T. S. Huang, "One-class SVM for Learning in Image Retrieval", *ICIP'2001*, Thessaloniki, Greece, October 7-10, 2001

[2]   I. J. Cox, M. Miller, T. Minka, P. Yianilos, "An Optimized Interaction Strategy for Bayesian Relevance Feedback", *CVPR'98,* Santa Barbara, CA. June 1998

[3]   P. Hong, Q. Tian, T. S. Huang, "Incorporate Support Vector Machines to Content-Based Image Retrieval with Relevance Feedback", *ICIP'2000*, Vancouver, Sep 10-13, 2000.

[4]   Y. Ishikawa, R. Subramanya, and C. Faloutsos, "Mind-Reader: Query databases through multiple examples", *24th VLDB Conf.* (New York), 1998

[5]   Jorma Laaksonen, M. Koskela, and E. Oja. "PicSOM: Self-Organizing Maps for Content-Based Image Retrieval," *Proc. IJCNN'99*. Washington, DC. July 1999

[6]   S. MacArthur, C. Brodley, and C. Shyu, "Relevance Feedback Decision Trees in Content-Based Image Retrieval," workshop *CBAIVL, CVPR'00*, June 12, 2000.

[7]   C. Meilhac and C. Nastar. "Relevance feedback and category search in image databases," In *IEEE Int'l Conf. on Multimedia Comp. and Sys.*, Florence, Italy, June 1999

[8]   S. Minka, G. Rätsch, J. Weston, B. Schölkopf, A. Smola, and K. Muller, "Fisher Discriminant Analysis with Kernels," in *IEEE workshop on Neural Networks for Signal Processing*, 1999

[9]   J. Peng, B. Bhanu, and S. Qing, "Probabilistic feature relevance learning for content-based image retrieval", *Computer Vision and Image Understanding*, 75:150-164, 1999

[10] R. W. Picard, T. P. Minka, and M. Szummer, "Modeling User Subjectivity in Image Libraries", *ICIP'96,* Lausanne, Sept. 1996.

[11] K. Porkaew, S. Mehrotra, and M. Ortega, "Query Reformulation for Content Based Multimedia Retrieval in MARS", *IEEE Int'l Conf. Multimedia Computing and Systems (ICMCS'99)*, June, 1999

[12] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval", *IEEE Tran Circuits and Systems for Video Technology*, Vol 8 No 5, 1998, 644-655

[13] Y. Rui, T. S. Huang, "Optimizing learning in image retrieval", *CVPR'00*, Hilton Head Island, SC, June 2000

[14] Gerald Salton, *Automatic text processing*, Addison-Wesley, Reading, Mass., 1989

[15] S. Santini, and Jain, R., "Integrated Browsing and Querying for Image Database," *IEEE Multimedia*, Vol. 7, No.3, 2000, page 26-39

[16] R. Schettini, G. Ciacca, and I. Gagliardi, "Content-based Color Image Retrieval with Relevance Feedback", *ICIP'99*, Kobe, 1999.

[17] Kinh Tieu and Paul Viola, "Boosting Image Retrieval", *CVPR'00*, Hilton Head, SC , 2000

[18] N. Vasconcelos, A. Lippman, "Bayesian relevance feedback for content-based image retrieval", *Workshop CBAIVL, CVPR'00*, Hilton Head Island, SC, 2000

[19] M. Worring, A. Smeulders, and S. Santini "Interaction in Content-based Image Retrieval: a state-of-the-art review" *Int'l Conf. on Visual Information Systems, Visual 2000*, Lyon, France, August 2000

[20] Y. Wu, Q. Tian, T. S. Huang, "Discriminant EM Algorithm with Application to Image Retrieval," *CVPR'00,* Hilton Head Island, South Carolina, June 13-15, 2000.

[21] X. S. Zhou, T. S. Hunag, "A generalized relevance feedback scheme for image retrieval", *Proc. of SPIE Vol. 4210: Internet Multimedia Management Systems*, Boston, MA. November 6-7, 2000

[22] X. S. Zhou, T. S. Huang, "Edge-based structural feature for content-based image retrieval", *Pattern Recognition Letters*, Vol 22/5, Apr. 2001. pp 457-468.