Feature Relevance Learning with Query Shifting for Content-Based Image Retrieval

Douglas R. Heisterkamp
doug@cs.okstate.edu

Jing Peng
jpeng@cs.okstate.edu

H. K. Dai
dai@cs.okstate.edu

Department of Computer Science
Oklahoma State University
Stillwater, OK 74078

Abstract

Probabilistic feature relevance learning (PFRL) is an effective technique for adaptively computing local feature relevance for content-based image retrieval. It however becomes less attractive in situations where all the input variables have the same local relevance, and yet retrieval performance might still be improved by simple query shifting. We propose a retrieval method that combines feature relevance learning and query shifting to try to achieve the best of both worlds. We use a linear discriminant analysis to compute the new query and exploit the local neighborhood structure centered at the new query by invoking PFRL. As a result, the modified neighborhoods at the new query tend to contain sample images that are more relevant to the input query. The efficacy of our method is validated using both synthetic and real world data.

1. Introduction

Probabilistic feature relevance learning for content-based image retrieval [8] computes flexible metrics for producing retrieval neighborhoods that are elongated along less relevant feature dimensions and constricted along most influential ones. The technique has shown promise in a number of image database applications. It, however, becomes less appealing in situations where all the input variables have the same local relevance, and yet retrieval performance might still be improved by simple query shifting.

On the other hand, MARS [9] is a simple query shifting mechanism that attempts to improve retrieval performance by adaptively moving the input query toward relevant retrievals and, at the same time, away from irrelevant ones. Similarity computation remains fixed throughout the retrieval process. While MARS has been shown to improve retrieval performance in simple tasks, it is clear that in many problems the mere shifting of the query is insufficient to achieve desired goals, as we shall see later.

In this paper, we propose a novel, principled approach that combines probabilistic feature relevance learning, as in [8], and query shifting, as in [9], to try to achieve the best of both worlds for content-based image retrieval [1, 6, 8, 9]. We use a linear discriminant analysis to compute the new query and exploit the local neighborhood structure centered at the new query by invoking PFRL. As a result, the modified neighborhoods at the new query tend to contain sample images that are more relevant to the input query.

2. Feature Relevance Learning with Query Shifting

We begin this section by briefly introducing the basic ideas behind PFRL and query shifting. We then describe in detail our method that combines PFRL and query shifting in a principled way.

2.1. PFRL

In PFRL [8], retrieved images with relevance feedback are used to compute local feature relevance. Let $\mathcal{R}_{\text{KNN}} = \{\mathbf{x}_j, y_j\}_1^K$ be the set of K retrievals, where $\mathbf{x}_j$ denotes the feature vector representing the jth retrieved image, and y_j is either 1 (relevant image) or 0 (irrelevant image) marked by the user as the class label associated with $\mathbf{x}_j$. If we let the class label $y \in \{0, 1\}$ at query $\mathbf{x}$ be treated as a random variable from a distribution with the probabilities $\{\Pr(1|\mathbf{x}), \Pr(0|\mathbf{x})\}$, we have

$$f(\mathbf{x}) \doteq \Pr(y = 1|\mathbf{x}) = E(y|\mathbf{x}).$$

In the absence of any variable assignments, the least-squares estimate for $f(\mathbf{x})$ is $E[f] = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x}$, where $p(\mathbf{x})$ is the joint density. Now given only that $\mathbf{x}$ is known at dimension $x_i = z_i$. The least-squares estimate becomes $E[f|x_i = z_i] = \int f(\mathbf{x})p(\mathbf{x}|x_i = z_i)d\mathbf{x}$. Here $p(\mathbf{x}|x_i = z_i)$ is the conditional density of the other input variables.

In image retrieval, $f(\mathbf{z}) = 1$, where $\mathbf{z}$ is the query. Then $[(f(\mathbf{z}) - 0) - (f(\mathbf{z}) - E[f|x_i = z_i])] = E[f|x_i = z_i]$ represents a reduction in error between the two predictions. Thus, a measure of feature relevance at query $\mathbf{z}$ can be defined as

$$r_i(\mathbf{z}) = E[f|x_i = z_i].$$

The relative relevance can be used as a weighting scheme for a weighted K-nearest neighbor search (KNN):

$$w_i(\mathbf{z}) = \exp(Tr_i(\mathbf{z}))/\sum_{l=1}^{q}\exp(Tr_l(\mathbf{z})).$$

Here T is a parameter that can be chosen to maximize (minimize) the influence of r_i on w_i. For further details, see [8].

2.2. Query Shifting

The *Standard Rocchio* equation is commonly used in the information retrieval field to determine the next query location based on relevance feedback [10]. The Standard Rocchio is

$$\mathbf{z}' = \alpha\mathbf{z} + \beta\frac{1}{n_{\mathrm{r}}}\sum_{\mathbf{x}\in\mathcal{R}_{\mathrm{r}}}\mathbf{x} - \gamma\frac{1}{n_{\mathrm{i}}}\sum_{\mathbf{x}\in\mathcal{R}_{\mathrm{i}}}\mathbf{x},$$

where $\mathbf{z}'$ is the new query location, $\mathbf{z}$ is the initial query location, $\mathcal{R}_{\mathrm{r}}$ is the set of relevant retrievals, $\mathcal{R}_{\mathrm{i}}$ is the set of irrelevant retrievals, n_{r} is the number of relevant retrievals, n_{i} is the number of irrelevant retrievals. The second term is equivalent to $\beta\boldsymbol{\mu}_{\mathrm{r}}$ where $\boldsymbol{\mu}_{\mathrm{r}}$ is the mean of the relevant retrievals. The third term is equivalent to $\gamma\boldsymbol{\mu}_{\mathrm{i}}$ where $\boldsymbol{\mu}_{\mathrm{i}}$ is the mean of the irrelevant retrievals. The Standard Rocchio expressed in terms of the retrieval means is $\mathbf{z}' = \alpha\mathbf{z} + \beta\boldsymbol{\mu}_{\mathrm{r}} + \gamma.\boldsymbol{\mu}_{\mathrm{i}}$ The values for the parameters α, β, and γ are determined by experimental runs over the database. The settings of $\alpha = 1.0, \beta = 0.75, \gamma = 0.25$ were reported in [10] to perform well in many cases. A common alternative is to ignore the influence of irrelevant retrievals ($\gamma = 0$) [3]. Another approach takes this theme further by setting $\alpha = 0$, $\beta = 1$, $\gamma = 0$, *i.e.*, the new query is $\boldsymbol{\mu}_{\mathrm{r}}$. Moving to $\boldsymbol{\mu}_{\mathrm{r}}$ has been claimed to be the optimal new query location [4]. It is easy to see that it is not since it ignores the effect of irrelevant retrievals [1]. See Figure 1, where moving to the positive mean also moves closer to the negative mean.

[1]It is optimal based on their criterion of minimum distance to relevant retrievals, but the criterion really should be a multiple criterion optimization of minimizing distance to relevant retrievals and maximizing distance to irrelevant retrievals.

2.3. Combining PFRL with Query Shifting

We present a hybrid system for learning feature relevance that seeks to draw upon the exploitation feature of PFRL and the exploration feature of query shifting.

For a given query image $\mathbf{z}$ in a q-dimensional feature space, we explore a new query $\mathbf{z}'$ in the q-dimensional feature space for more relevant retrievals, if necessary, as follows. The computation of $\mathbf{z}'$ is aided by a feature extraction that transforms from the q-dimensional feature space to a one-dimensional space, which retains sufficient information of the retrieval images.

Classical discriminant analysis (see, for example, [11]) attempts to project patterns into a space with lower dimensionality than the original pattern space. The discriminant analysis projection maximizes the inter-class scatter while keeping the intra-class scatter constant. When the number of pattern classes is two, like in our case, the discriminant analysis projection can be realized by the one-dimensional Fisher linear discriminant projection, which requires to calculate the intra-class and inter-class scatter matrices.

To determine the exact location of $\mathbf{z}'$ without computing the scatter matrices, we consider the projections of all K retrieved images onto the line L passing through the two sample means $\boldsymbol{\mu}_{\mathrm{r}}$ and $\boldsymbol{\mu}_{\mathrm{i}}$. Parameterize the points on L as $L(\lambda) = \lambda\boldsymbol{\mu}_{\mathrm{r}} + (1-\lambda)\boldsymbol{\mu}_{\mathrm{i}}$ with real λ. We will let $\mathbf{z}' = L(\lambda^*)$ for some suitably chosen λ^*.

Our objective is to find λ^* such that in the vicinity of $L(\lambda^*)$, the frequency of retrieving relevant class-1 images is high. This suggests to select λ^* that maximizes the conditional expectation of an image pattern $\mathbf{x}$ given that the component of $\mathbf{x}$ in the direction of L is $L(\lambda^*)$. That is,

$$\lambda^* = \arg\max_{\lambda} E[f(\mathbf{x}) \mid \mathrm{proj}_L(\mathbf{x}) = L(\lambda^*)].$$

After computing $\mathbf{z}'$, we exploit the neighborhood structure centered at the next query $\mathbf{z}'$ by invoking PFRL on all previous (cumulative) retrieval images to generate the relative relevance weights used to determine the KNN in the next iteration.

An estimate for the conditional expectation for a point on L can be determined by projecting the KNN retrievals onto L and using the following equation [2, 8]

$$\hat{E}[f(\mathbf{x})|\mathrm{proj}_L(\mathbf{x}) = L(\lambda)] = \frac{\sum_{\mathbf{x}\in\mathcal{R}_{\mathrm{r}}^{+}} y(\mathbf{x})1(|\mathrm{proj}_L(\mathbf{x}) - L(\lambda)| \leq \Omega)}{\sum_{\mathbf{x}\in\mathcal{R}_{\mathrm{KNN}}^{+}} 1(|\mathrm{proj}_L(\mathbf{x}) - L(\lambda)| \leq \Omega)},$$

where $1(\cdot)$ is an indicator function for its predicate argument, $y(\mathbf{x})$ is the label of a retrieved image $\mathbf{x}$, and $\mathcal{R}_{\mathrm{KNN}}^{+}$ and $\mathcal{R}_{\mathrm{r}}^{+}$ are the sets of cumulated retrieved images and retrieved relevant images respectively. The retrieved images are cumulated for each individual query sequence. When the user initiates a query with a new image, the accumulations are reset.

The value of Ω is chosen such that

$$\sum_{\mathbf{x} \in \mathcal{R}_{\mathrm{kNN}}} 1(|\mathrm{proj}_L(\mathbf{x}) - L(\lambda)| \leq \Omega) = C.$$

Due to the discrete nature of the estimate, a segment of L will maximize $\hat{E}$, (possibly multiple segments, in which case we choose the largest segment). Any of the points on the segment maximizing $\hat{E}$ can be chosen for the next query location. We chose the mean of the relevant samples that contributed to the estimate, $\hat{E}$. With this choice, moving to the relevant mean (i.e., $\lambda = 1$) is obtained as a special case by letting $C = K$. The α, β, γ parameters are determined from λ by setting $\alpha = 0$, $\beta = \lambda$, and $\gamma = \lambda - 1$.

3. Experimental Results

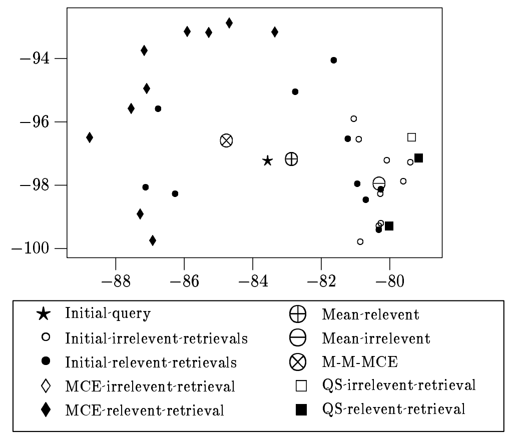
In the following we compare the retrieval methods of query shifting, PFRL, and PFRL combined with query shifting on synthetic and real data. We also compare $\boldsymbol{\mu}_{\mathrm{r}}$ and $L(\lambda^*)$ for the query shift location. In all the experiments in this section, the data is normalized along each feature dimension of each entire data set.

First an example query using synthetic data is presented. Then the average retrieval precision of the different methods on real data is presented.

The synthetic data used is the 2D example data that was provided with the *Multivariate Data Generation Software* [5] from the ICPR 2000 Algorithm Performance Contest. A query (represent as $\star$) at location (-83.5, -97.2) is presented in Figure 1 with both the shift to $\boldsymbol{\mu}_{\mathrm{r}}$ ($\oplus$ in figure) and the shift to $L(\lambda^*)$ ($\otimes$ in figure). Just query shifting is used in this example. The shift to $\boldsymbol{\mu}_{\mathrm{r}}$ also moves closer to $\boldsymbol{\mu}_{\mathrm{i}}$ ($\ominus$ in figure) with the resulting effect on the KNN of replacing three relevant images with two relevant and one irrelevant. The shift to $L(\lambda^*)$ moves away from $\boldsymbol{\mu}_{\mathrm{i}}$ with the resulting effect on the KNN by replacing three relevant and seven irrelevant images with ten relevant images.

The retrieval methods of PFRL, shifting to $\boldsymbol{\mu}_{\mathrm{r}}$, shifting to $L(\lambda^*)$, PFRL+$\boldsymbol{\mu}_{\mathrm{r}}$, and PFRL+$L(\lambda^*)$ was applied to each of the following four databases. For all of the retrieval methods, each image in the database was selected as a query. For each iteration, upto five iterations, the 20 nearest neighbors were returned with relevance feedback. The average retrieval precision for each method and each database is presented in Figure 3.

Database 1. The data (Texture Data) was obtained from MIT Media Lab at: whitechapel.media.mit.edu/pub/VisTex. There are a total of 640 images of 128×128 in the database with 15 classes. The images in this database are represented by 8 Gabor filters (2 scales and 4 orientations). Examples of the textures are presented in Figure 2.



Legend

Figure 1. Synthetic 2D data: example query

Figure 2. Example images from texture data

Database 2. This is a set (Sonar Data), also taken from [7], of 208 data points having two classes (Mines and Rocks) with equal number of instances in each class. The data are represented by 60 features. For details, see [7].

Database 3. This data set (Vowel Data) has $q = 10$ measurements and 11 classes. There are a total of $N = 528$ samples in this database. This set is also taken from [7].

Database 4. The data set (Segmentation Data), taken from the UCI repository [7], consists of images that were drawn randomly from a database of 7 outdoor images. The images were hand-segmented by the creators of the database to classify each pixel. Each image is a region. There are 7 classes, each of which has 330 instances. Thus, there are a total of 2310 images in the database. These images are represented by 19 real valued attributes.

The retrieval precision of PFRL combined with query shifting consistently outperformed just query shifting and PFRL individually. Shifting to $L(\lambda^*)$ outperformed shifting to $\boldsymbol{\mu}_{\mathrm{r}}$ both individually and when combined with PFRL, though the magnitude of improvement is much less than the magnitude of improvement of either combined method over the individual methods.

The close performance of shifting to $L(\lambda^*)$ and shifting to $\boldsymbol{\mu}_{\mathrm{r}}$ can be explained by noting that often $L(\lambda^*)$ is at or very close to $\boldsymbol{\mu}_{\mathrm{r}}$. For example using the texture data, 84% of the time the maximum condition expectation is very close to $\boldsymbol{\mu}_{\mathrm{r}}$ (within 0.01 of the mean-mean distance). Only 3% of the time is it far away (greater than 0.3 of the mean-

mean distance). Thus shifting to $L(\lambda^*)$ performs the same as shifting to $\boldsymbol{\mu}_r$in the common case and in the infrequent case that $\boldsymbol{\mu}_r$ is a bad location, it performs much better.

4. Summary

This paper presents a novel method that combines probabilistic feature relevance learning and query shifting to try to achieve the best of both worlds. This method uses a linear discriminant analysis to compute the new query upon which to estimate local retrieval neighborhood using PFRL. As a result, the modified neighborhood at the new query tends to contain data samples that are more relevant to the input query. The experimental results using both synthetic and real data show convincingly that feature relevance learning coupled with query shifting outperformed either PFRL or query shifting alone.

A potential extension to the technique described in this paper is to consider additional derived variables (features) for local relevance estimate and query shifting, thereby contributing to the overall retrieval performance. The challenge is to be able to have a mechanism that computes such informative derived features efficiently.

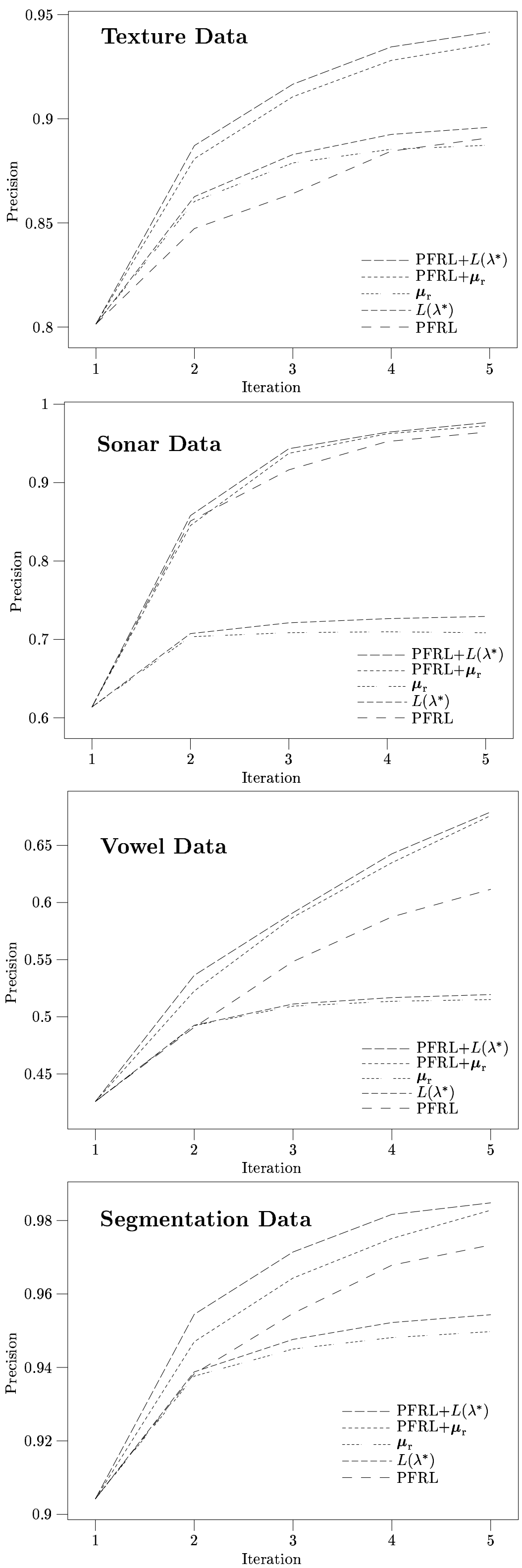


Figure 3. Precision graphs

References

[1] M. F. *et al*. Query by image and video content: The qbic system. *IEEE Computer*, pages 23–31, September 1995.

[2] J. H. Friedman. Flexible metric nearest neighbor classification. Technical report, Department of Statistics, Stanford University, 1994.

[3] D. A. Grossman and O. Frieder. *Information Retrieval: Alogithms and Heuristics*. Kluwer, 1999.

[4] Y. Ishikawa, R. Subramanya, and C. Faloutsos. Mindreader: Querying databases through multiple examples. Technical Report CMU-CS-98-119, Carnegie Mellon University, 1998.

[5] J. Liang. Multivariate data generate software. http://isl.ee.washington.edu/IAPR/ICPR00/gendata/ packages/gendata.tar.gz.

[6] T. Minka and R. Picard. Interactive learning with a 'society of models. *Pattern Recognition*, 30(4):565–81, April 1997.

[7] P. Murphy and D. Aha. Uci repository of machine learning databases. www.cs.uci.edu/∼mlearn/MLRepository.html.

[8] J. Peng, B. Bhanu, and S. Qing. Probabilistic feature relevance learning for content-based image retrieval. *Computer Vision and Image Understanding*, 75(1/2):150–164, 1999.

[9] Y. Rui, T. Huang, and S. Mehrotra. Content-based image retrieval with relevance feedback in mars. In *Proceedings of IEEE International Conference on Image Processing*, pages 815–818, October 1997.

[10] P. Schauble. *Multimedia Information Retrieval: Content-Based Information Retrieval from Large Text and Audio Databases*. Kluwer, Boston, 1997.

[11] S. S. Wilks. *Mathematical Statistics*. John Wiley & Sons, Inc., New York, 1963.