Nonparametric Discriminant Analysis in Relevance Feedback for Content-based Image Retrieval

Dacheng Tao and Xiaoou Tang
Department of Information Engineering
The Chinese University of Hong Kong
{dctao2, xtang}@ie.cuhk.edu.hk

Abstract

Relevance feedback (RF) has been widely used to improve the performance of content-based image retrieval (CBIR). How to select a subset of features from a large-scale feature pool and to construct a suitable dissimilarity measure are key steps in RF. Biased discriminant analysis (BDA) has been proposed to select features during relevance feedback iterations. However, BDA assumes all positive feedbacks form a single Gaussian distribution which may not be the case for CBIR. Although kernel BDA can overcome the drawback to some extent, the kernel parameter tuning makes the online learning unfeasible. To avoid the parameter tuning problem and the single Gaussian distribution assumption in BDA, we construct a new nonparametric discriminant analysis (NDA). To address the small sample size problem in NDA, we introduce the regularization method and the null-space method. Because the regularization method may meet the ill-posed problem and the null-space method will lose some discriminant information, we propose here a full-space method. The proposed full-space NDA is demonstrated to outperform BDA based RF significantly based on a large number of experiments in Corel database with 17,800 images.

1. Introduction

Relevance feedback (RF) [1] is an important tool to improve the performance of content-based image retrieval (CBIR) [2]. In a RF process, the user first labels a number of relevant retrieval results as positive feedbacks and some irrelevant retrieval results as negative feedbacks. Then the system refines all retrieval results based on these feedbacks. The two steps are carried out iteratively to improve the performance of image retrieval system by gradually learning the user’s perception.

Many RF methods have been developed in recent years. One approach [1] adjusts the weights of various features to adapt to the user’s perception. Another approach [3] estimates the density of the positive feedback examples. Support vector machine (SVM) has also been used as a classification method for RF [4]. These methods all have their own limitations. The method in [1] is only heuristic based. The density estimation method [3] loses information contained in negative samples. Classification based method in [4] treats the positive and negative samples equally.

Recently, biased discriminant analysis (BDA) [5,6,7] has been used as a feature selection method to improve RF, because BDA models the RF better than many other methods. However, BDA assumes all positive samples form a single Gaussian distribution, which means all positive samples should be similar with similar view angle, similar illumination, etc. Clearly, this is not the case for CBIR. The kernel-based learning is used in BDA to overcome the problem. However, kernel-based learning has to rely on parameter tuning, which makes the online learning unfeasible.

To avoid the parameter tuning problem and the single Gaussian distribution assumption in BDA, we develop a new discriminant analysis using a nonparametric approach. The proposed nonparametric discriminant analysis (NDA) has the following properties: 1. NDA assumes all positive samples are alike and each negative sample is negative in its own way; 2. NDA does not require all positive samples form a single Gaussian distribution. 3. NDA, similar to BDA and KBDA, may meet the Small-Sample-Size (SSS) problem. In this paper, we will solve the SSS problem with three methods: 1. the regularization method, which is used by Zhou in BDA [5]; 2. the null-space method [8], which is a popular method to solve the SSS problem in linear discriminant analysis for face recognition; 3. the full-space method, which is proposed to preserve all discriminant information of NDA.

2. Nonparametric discriminant analysis

To better understand the proposed NDA based RF schemes, we first give a brief review of BDA.

2.1. BDA

BDA [5] tries to find the subspace to discriminate the positive (the only class concerned by the user) and negative samples (unknown number of classes). It is spanned by a set of vectors \( W \) maximizing the ratio between the positive covariance matrix \( S_p \) and the biased matrix \( S_v \).
\[ W = \text{arg} \max_w \frac{|w^T S_w w|}{|w^T S_b w|}. \]  

Let the training set contains \(N_x\) positive and \(N_y\) negative samples. Then \(S_x\) and \(S_y\) are defined as,

\[
\begin{align*}
S_x &= \sum_{i=1}^{N_x} (x_i - \bar{x})(x_i - \bar{x})^T, \\
S_y &= \sum_{i=1}^{N_y} (y_i - \bar{y})(y_i - \bar{y})^T
\end{align*}
\]

where \(x_i\) denote the positive samples, \(y_i\) denote the negative samples, \(\bar{x}\) is the mean vector of the positive samples, and \(\bar{y}\) can be computed from the eigenvectors of \(S_y\). Firstly, BDA minimize the distance between all positive feedbacks and all negative feedbacks. Then BDA maximize the variance of the positive samples. Then BDA maximize \(\text{max arg} \frac{1}{N_y} \sum_{i=1}^{N_y} \|y_i - \bar{y}\|^2\) over the null-space method, and the new full-space method.

\[ S_x' = (1 - \gamma) S_x + \frac{\gamma}{n} \text{tr} \{S_x\} I \\
S_y' = (1 - \gamma) S_y + \frac{\gamma}{n} \text{tr} \{S_y\} I \]

where \(\gamma\) and \(\mu\) control the shrinkage toward a multiple of the identity matrix. \(\text{tr} \{\cdot\}\) is the trace operation.

It is well known that regularization method may meet the ill-posed issue. Hence, we select the null-space to overcome the ill-posed issue.

**3.2. Null-space method**

Null-space linear discriminant analysis (LDA) [8] accepts high-dimensional data as the input, and optimizes LDA in the null space of within class scatter matrix. Here, we generalize the null-space idea for NDA. The null space of \(S_x\) is first calculated as:

\[ Y^T S_x Y = 0 \]

where \(Y\) are eigenvectors with zero eigenvalues and \(Y^T Y = I\). \(S_x\) is projected onto the null space of \(S_x\):

\[ \tilde{S}_x = Y^T S_x Y \]

The eigenvectors \(U\) of \(\tilde{S}_x\) with largest eigenvalues are selected to form the transformation matrix as:

\[ W = YU \]

**3.3. Full-space method**

Null-space method loses the information in the principle space of the within class scatter matrix. In order to preserve all discriminant information, we compute features from both the null space and the principle space of \(S_x\), and then integrate the two parts with a suitable weighting. A rational choice of the weighting is to select
a small eigenvalue of $S_x$. The algorithm first computes
the eigenvalues of $\tilde{S}_y$ as,

$$Y' \tilde{S}_y Y = D_y,$$

where $D_y = \text{diag}(\lambda_1, \ldots, \lambda_l, \ldots, \lambda_k, \ldots, \lambda_m)$, $\lambda_m = \varepsilon \lambda_1$, and $\varepsilon$
is a user selected threshold value (such as 0.01).

For a given $\varepsilon$, the eigenvalue matrix $D_y$ is replaced by $D_y = \text{diag}(\lambda_1, \ldots, \lambda_l, \ldots, \lambda_k, \ldots, \lambda_m)$. All values, which are smaller than $\lambda_m$, are substituted by $\lambda_m$. After the substitution, $\tilde{S}_y$ is projected onto the space by:

$$\hat{S}_y = D_y^{-1/2} Y' \hat{S}_y Y D_y^{-1/2}.$$

Finally, the eigenvectors $U$ of $\hat{S}_y$ with largest eigenvalues are selected to form the transformation matrix,

$$W = Y D_y^{-1/2} U D_y^{-1/2}.$$

4. Experimental results

In our experiments, three main features, color, texture, and shape are extracted and used to represent the corresponding image. For color feature, we use the color histogram [9] in HSV color space. Here, the color histogram is quantized into 256 levels. Hue, Saturation, and Value are quantized into 8, 8, and 4 bins respectively. Texture is extracted from Y component in YCrCb space by pyramidal wavelet transform (PWT) with Haar wavelet. The mean value and standard deviation are calculated for each sub-band at each decomposition level. The feature length is $2 \times 4 \times 3$. For shape feature, edge histogram [10] is calculated on Y component in YCrCb color space. Edges are grouped into four categories, horizontal, 45 diagonal, vertical, and 135 diagonal. We combine the color, texture, and shape features into a feature vector, and then we normalize each feature to a normal distribution.

In this part, a large number of statistical experiments are performed based on a subset of the Corel Photo Gallery, which includes 17,800 images with 90 concepts (relabeled by ourselves). The experiments are simulated by the computer automatically. First, 300 queries are randomly selected from the data, and then RF is done by computer as: top 5 query relevant and irrelevant images are marked as positive and negative feedbacks in the top 48 images, respectively.

In this paper, precision and standard deviation (SD) are used to evaluate the performance of a RF algorithm. Precision is the ratio of the number of relevant images retrieved to the top $N$ retrieved images. Precision curve is the averaged precision values of the 300 queries, and SD curve is the SD values of 300 queries’ precision. The precision curve evaluates the effectiveness of a given algorithm and SD curve evaluates the robustness of the algorithm. In precision and SD curves, the total feedback times are 9, with 0 feedback referring to the retrieval based on Euclidean distance measure without RF.

4.1. K nearest neighbor evaluation

The experiment shows NDA is insensitive to the $k$ value of the k-nearest-neighbor. Figure 1, 2, and 3 show the top 30 retrieved results with 3, 6, and 9 feedback iterations by the regularization method, null-space method, and full-space method, respectively. Because all curves are flat, we can draw the conclusion that NDA is insensitive to the $k$ value in $k$ nearest neighbor.

### 4.2. Small samples size problem

Fig. 4 shows the performance of the full-space method, the null-space method, and the regularization method in NDA to solve the SSS problem. From the left subfigure in Fig. 4, we can see the precision curve of full-space method is higher than that of null-space method and regularization method, meanwhile the SD curve of full-space method is lower than that of null-space method and regularization method. Hence we can draw the conclusion the new full-space method can work better than the existing null-space method and regularization method. Meanwhile, the null-space method can outperform the regularization method.
4.3. Evaluation experiments

We will compare the new full-space NDA with the existing state-of-the-art algorithms, which are BDA [5], SVM [4], and constrained SVM (CSM) [11]. Results in Fig. 5 shows that the full-space NDA by 3-nearest-neighbor can significantly improve the CBIR RF compared with all the other algorithms [4,5,11].

5. Conclusion

In this paper, we propose a new nonparametric discriminant analysis (NDA) for relevance feedback (RF) in content-based image retrieval. To address the small sample size problem in NDA, we proposed a full-space method. Based on a large number of experiments with 17,800 images, we can draw the conclusion that the new full-space NDA based RF can work much better than the state-of-the-art methods.

6. Acknowledgement

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong SAR. (Project no. AoE/E-01/99).

7. References