

Tensor Graph-optimized Linear Discriminant Analysis

Jianjun Chen
Department of Information and Electron of Yuanpei of College
Shaoxing University Shaoxing
Shaoxing 312000
China
Chengjianjun_2012@163.com



ABSTRACT: Graph-based Fisher Analysis (GbFA) is proposed recently for dimensionality reduction, which has the powerful discriminant ability. However, GbFA is based on the matrix-to-vector way, which not only costs much but also loses spatial relations of pixels in images. Therefore, Tensor Graph-based Linear Discriminant Analysis (TGbLDA) is proposed in the paper. TGbLDA regards samples as data in tensor space and gets projection matrixes through the iteration way. Besides, TGbLDA inherits merits of GbFA. Experiments on Yale and YaleB face datasets demonstrate the effectiveness of our proposed algorithm.

Categories and Subject Descriptors:

I.7.5 Graphics Recognition and Interpretation, G.2.2 Graph Theory

General Terms: Graph Analysis, Linear Discriminant Analysis, Graph Algorithms

Keywords: Dimensionality Reduction, Graph-based Fisher Discriminant Analysis, Tensor Data, Face Recognition

Received: 3 July 2013, Revised 14 August 2013, Accepted 19 August 2013

1. Introduction

High-dimensional feature of data not only increase the computational complexity and contain plenty of redundant information, therefore dimensionality reduction is essential the processing step for data mining. The destination of dimensionality reduction is to map high-dimensional data

into high-dimensional data without losing certain feature as little as possible. Linear Discriminant Analysis (LDA) [1] or Fisher Discriminant Analysis (FDA) [2] is common dimensionality reduction method. LDA or FDA attempts to preserve the separability within classes as much as possible and has good discriminant performance. There are more and more attentions on LDA or FDA from researchers. Representative algorithms include Pseudo-inverse Linear Discriminant Analysis (PLDA) [3], regular Linear Discriminant Analysis (RLDA) [4], Penalized Discriminant Analysis (PDA) [5], LDA/GSVD [6], LDA/QR [7], Orthogonal Linear Discriminant Analysis (OLDA) [8], Null Space Linear Discriminant Analysis (NLDA) [9], Direct Linear Discriminant Analysis (DLDA) [10], Nonparametric Discriminant Analysis (NDA) [11], Local Fisher Discriminant Analysis (LFDA) [12], Multi-label Linear Discriminant Analysis (MLDA) [13] and Local Linear Discriminant Analysis (LLDA) [14]. However the effectiveness of these algorithms is still limited because the number of the available projection directions is lower than the class number. Moreover, these algorithms are proposed based on the data approximately obeying a Gaussian distribution, which cannot always be satisfied in the real-world applications. Inspired by the success of the graph-based embedding dimensionality reduction techniques, Cai et al [15] proposes a novel supervised dimensionality reduction algorithm, called the Graph-based Fisher Analysis (GbFA). GbFA don't need to obey a Gaussian distribution. GbFA redefined The intrinsic graph based on the same-class samples and the penalty graph based on the not-same-class samples. GbFA will make the original neighbor same-class samples much closer in the output space while pushing apart the original

neighbor not-same-class samples in the output space, so GbFA encodes the discriminating information. However, GbFA need transform two or more dimensional feature matrix into feature vector, which leads to loss of spatial relation on pixels in face images. So researchers proposed tensor versions of LDA. He et al [16] proposed Tensor Linear Discriminative Analysis (TLDA). On the basic of LFDA, Zhao et al [17] proposed Tensor Locally Linear Discriminative Analysis, TLLDA).

Inspired by above analyses, a dimensionality reduction algorithm called Tensor Graph-based Linear Discriminant Analysis (TGBLDA) is proposed in the paper. TGBLDA regards two-dimensional face image as second-order tensor data and get two projections through the iteration loop with TGBLDA. Projected data not only preserve graph-based discriminant information but also preserve the spatial relations of pixels in face images. Experiment on Yale and YaleB demonstrate that our algorithm is efficient.

2. Graph-based Fisher Analysis (GbFA)

The objective function of GbFA is gotten as follows:

(1) Firstly, according to the theory of graph optimization, GbFA construct an intrinsic graph $G^c = \{X, W^c\}$ and a penalty graph $G^p = \{X, W^p\}$ as follows

$$W_{ij}^c = \begin{cases} \exp\{-\|x_i - x_j\|^2/t\} & \text{if } y_i = y_j \\ 0 & \text{if } y_i \neq y_j \end{cases} \quad (1)$$

$$W_{ij}^p = \begin{cases} \exp\{-\|x_i - x_j\|^2/t\} & \text{if } y_i \neq y_j \\ 0 & \text{if } y_i = y_j \end{cases} \quad (2)$$

where $t \in R$, W_{ij}^c indicates the importance degree of x_i and x_j in the same class. W_{ij}^p indicates the importance degree of x_i and x_j in the not same class.

(2) Secondly, with the embedding map T , the intra-class compactness can be characterized from the intrinsic graph, GbFA defines the square of the norm in the form of the matrix trace as follows:

$$G^c = \sum_{i=1}^n \sum_{j=1}^n \|y_i - y_j\|^2 W_{ij}^c \sum_{i=1}^n \sum_{j=1}^n \text{tr}\{T^T(x_i - x_j)(x_i - x_j)^T T\} W_{ij}^c \quad (3)$$

$$G^p = \sum_{i=1}^n \sum_{j=1}^n \|y_i - y_j\|^2 W_{ij}^p \sum_{i=1}^n \sum_{j=1}^n \text{tr}\{T^T(x_i - x_j)(x_i - x_j)^T T\} W_{ij}^p \quad (4)$$

Equation (3) can be further transformed into

$$G^c = \sum_{i=1}^n \sum_{j=1}^n \text{tr}\{T^T(x_i - x_j)(x_i - x_j)^T T\} W_{ij}^c \quad (5)$$

$$= \text{tr}\{T^T(2XDX - 2XW^cX)T\}$$

$$= 2\text{tr}\{T^T X(D^c - W^c)XT\}$$

$$= 2\text{tr}\{T^T X L^c XT\}$$

where $D^c = \text{diag}(D_{11}, \dots, D_{nn})$, $D_{ii} = \sum_{j=1}^n W_{ij}^c$ ($i = 1, \dots, n$), $L^c = D^c - W^c$ denotes the laplacian matrix.

Similarly, Equation (4) may be transformed into

$$G^p = 2\text{tr}\{T^T X L^p XT\} \quad (6)$$

(3)The objective function of GbFA is gotten as follows

$$\max_T \left(\frac{G^p}{G^c} \right) = \max_T \left(\frac{\text{tr}\{T^T X L^p XT\}}{\text{tr}\{T^T X L^c XT\}} \right) \quad (7)$$

3. Tensor Graph-based Linear Discriminant Analysis (TGBLDA)

3.1 Basic ideal

Given samples $X = \{x_i | x_j \in R^{n_1 \times n_2}, 1 \leq i \leq n\}$, X is regarded as one datasets in the second-order tensor space $R^{n_1} \otimes R^{n_2}$. The destination of TGBLDA is to find two projection matrixes $U(n_1 \times l_1)$ and

$U(n_2 \times l_2)$ to get $Y = U^T X V = \{y_i | y_j \in R^{l_1 \times l_2}, 1 \leq i \leq n, l_1 < n_1, l_2 < n_2\}$

(1) According to $\|A\|_F^2 = \text{tr}(AA^T)$, we can get

$$\max_{U, V} \left(\frac{\sum_{i=1}^n \sum_{j=1}^n \|y_i - y_j\|^2 W_{ij}^p}{\sum_{i=1}^n \sum_{j=1}^n \|y_i - y_j\|^2 W_{ij}^c} \right)$$

$$= \max_{U, V} \left(\frac{\sum_{i=1}^n \sum_{j=1}^n \text{tr}(y_i - y_j)(y_i - y_j)^T W_{ij}^p}{\sum_{i=1}^n \sum_{j=1}^n \text{tr}(y_i - y_j)(y_i - y_j)^T W_{ij}^c} \right) \quad (8)$$

$$= \max_{U, V} \left(\frac{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n U^T (x_i V - x_j V) (x_i V - x_j V)^T U W_{ij}^p \right)}{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n U^T (x_i V - x_j V) (x_i V - x_j V)^T U W_{ij}^c \right)} \right)$$

$$= \max_{U, V} \left(\frac{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n U^T (DV_i - DV_j) (DV_i - DV_j)^T U W_{ij}^p \right)}{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n U^T (DV_i - DV_j) (DV_i - DV_j)^T U W_{ij}^c \right)} \right)$$

where $DV_i = x_i V$

(2) According to $\|A\|_F^2 = \text{tr}(A^T A)$, we can get

$$\begin{aligned}
& \max_{U, V} \left(\frac{\sum_{i=1}^n \sum_{j=1}^n \|y_i - y_j\|^2 W_{ij}^p}{\sum_{i=1}^n \sum_{j=1}^n \|y_i - y_j\|^2 W_{ij}^c} \right) \\
& = \max_{U, V} \left(\frac{\sum_{i=1}^n \sum_{j=1}^n \text{tr} (y_i - y_j)^T (y_i - y_j) W_{ij}^p}{\sum_{i=1}^n \sum_{j=1}^n \text{tr} (y_i - y_j)^T (y_i - y_j) W_{ij}^c} \right) \quad (9) \\
& = \max_{U, V} \left(\frac{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n V^T (x_i - x_j)^T U U^T (x_i - x_j) V W_{ij}^p \right)}{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n V^T (x_i - x_j)^T U U^T (x_i - x_j) V W_{ij}^c \right)} \right) \\
& = \max_{U, V} \left(\frac{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n V^T (DU_i - DU_j)^T (DU_i - DU_j) V W_{ij}^p \right)}{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n V^T (DU_i - DU_j)^T (DU_i - DU_j) V W_{ij}^c \right)} \right)
\end{aligned}$$

where $DU_i = V^T x_i$

(3) The objective function of TGbLDA is listed as follows:

$$\begin{aligned}
& \max_{U, V} \left(\frac{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n U^T (DV_i - DV_j) (DV_i - DV_j)^T U W_{ij}^p \right)}{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n U^T (DV_i - DV_j) (DV_i - DV_j)^T U W_{ij}^c \right)} \right) \\
& = \max_{U, V} \left(\frac{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n V^T (DU_i - DU_j)^T (DU_i - DU_j) V W_{ij}^p \right)}{\text{tr} \left(\sum_{i=1}^n \sum_{j=1}^n V^T (DU_i - DU_j)^T (DU_i - DU_j) V W_{ij}^c \right)} \right) \quad (10) \\
& \quad s. t \sum_i \|X_i \times_1 U \times_2 V\|_F^2 = 1
\end{aligned}$$

3.2 Algorithm steps

Given $X = \{x_j \in R^{n_1 \times n_2}, 1 \leq j \leq n\}$, error ε , two projection matrixes $U (n_1 \times l_1)$ and $V (n_2 \times l_2)$ are gotten as follows:

(1) Set $U_{(1)} = I(n_1, l_1)$ and $V_{(1)} = I(n_2, l_2)$ where $U_{(1)}$ and $V_{(1)}$

denote unit matrix,

$X^V = \{x_i^u | x_i^u \in R^{l_1 \times l_2}, 1 \leq i \leq n\}$, $X^u = \{x_i^u | x_i^u \in R^{l_2 \times l_1}, 1 \leq i \leq n\}$

(2) Set loop variable $t=1$.

(3) Calculate $DV_i = x_i V_{(t)} (1 \leq i \leq n)$.

(4) According to Equation (8), calculate $U_{(t)}$ using gneralized matrix solution.

(5) Calculate $DU_i = x_i U_{(t)} (1 \leq i \leq n)$.

(6) According to Equation (9), calculate $V_{(t)}$ using gneralized matrix solution.

(7) If $\|U_{(t)} - U_{(t-1)}\| < \varepsilon$ and $\|V_{(t)} - V_{(t-1)}\| < \varepsilon$, then jump to (8), else $t = t + 1$ and jump to (4).

(8) Attain projection matrixes $U = U_{(t)}$ and $V = V_{(t)}$.

4. Experiments

4.1 Experimental datasets

In the experiment, Yale and YaleB face datasets are used and are described as follows:

(1) Yale contains 165 face images of 15 individuals. There are 11 images per subject, and these 11 images are, respectively, under the following different facial expression or configuration: center-light, wearing glasses, happy, left-light, wearing no glasses, normal, right-light, sad, sleepy, surprised and wink. In our experiment, the images are cropped to a size of 30×30 . Figure 1 shows a group of images in Yale.

(2) YaleB contains 2414 front-view face images of 38 individuals. For each individual, about 64 pictures were taken under various laboratory-controlled lighting conditions. In our experiments, we use the cropped images with the resolution of 30×30 . Figure 2 shows a group of images in YaleB.



Figure 1. A group of images in Yale



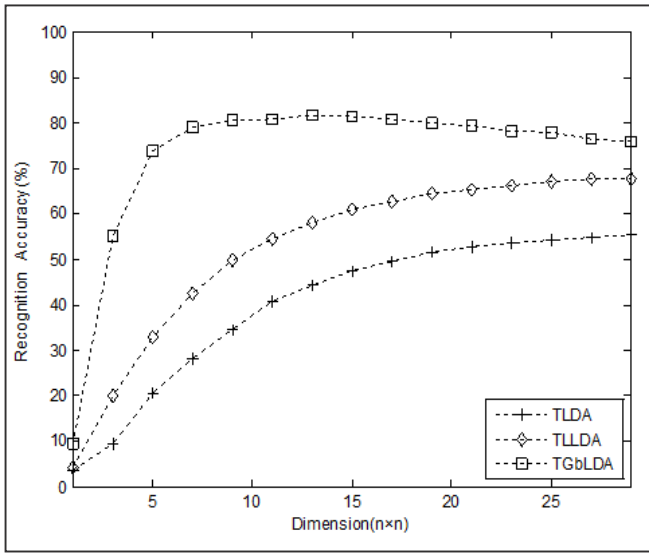
Figure 2. A group of images in Yale B

4.2 Experimental settings

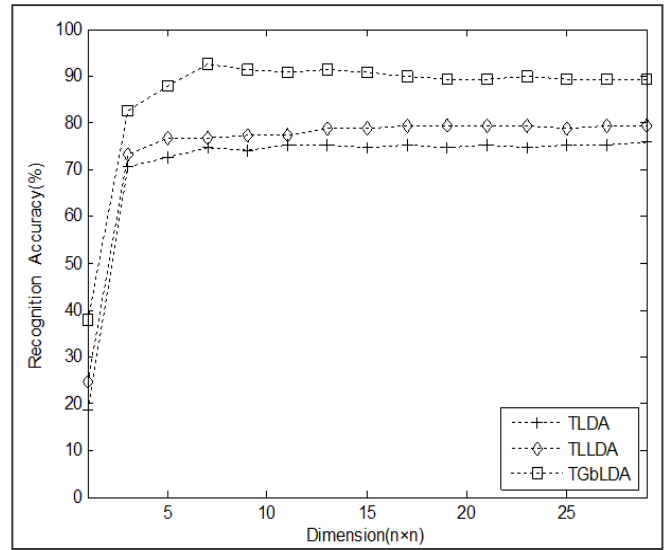
TLDA [15] and TLLDA [16] are popular tensor dimensionality reduction methods on LDA and have been successfully applied to face recognition. In the experiment, we test the performance of the two algorithms and the proposed algorithm on Yale and YaleB face datasets. Parameter settings of them are listed in Table 1.

Algorithms	Parameter settings
TLDA	no
TLLDA	no
TGbLDA	$k = 7, t = 1$

Table 1. Parameter settings of algorithm

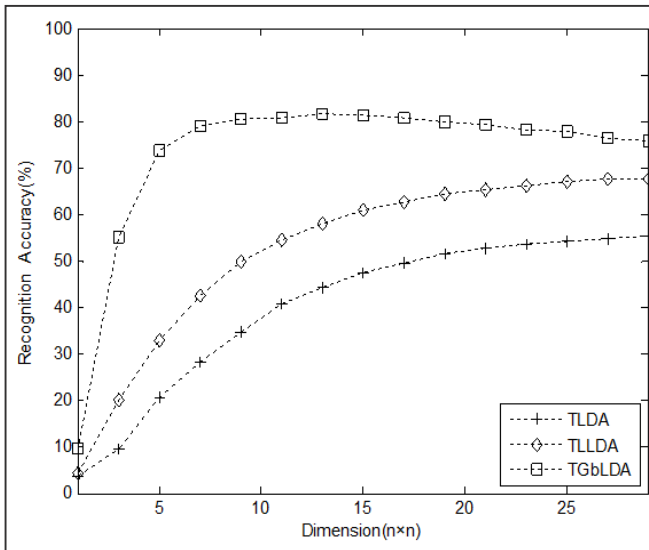


(a) $T = 3$

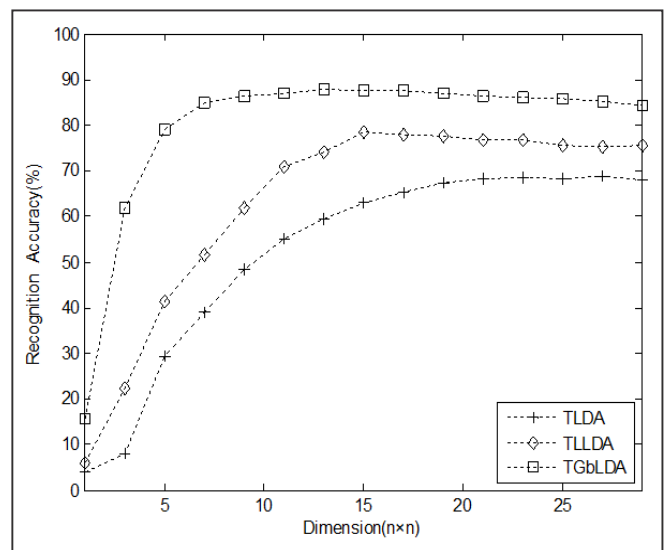


(b) $T = 6$

Figure 3. The recognition rates with different T on Yale



(a) $T = 10$



(b) $T = 20$

Figure 4. The recognition rates with different T on YaleB

Moreover, for Yale and YaleB face datasets, we randomly select T images per class for training and the remaining for test.

4.3 Experimental results and analysis

To verify the effectiveness of the proposed method, we evaluate the performance of the proposed method and compare these methods using the simplest nearest neighbor classifier. The recognition rates are shown in Figure 3 and Figure 4.

From above Figure 3 and Figure 4, we can draw following conclusions:

(1) With increases of the number of the dimension, the classification performance of all algorithms constantly increased. But when the subspace dimension attains a small threshold, performance of TGBLDA can obtain maximum performance. TLDA and TLLDA need higher

dimension for the maximum performance than TGBLDA .

(2) In contrast to TLDA, TGBLDA outperforms obviously, especially on YaleB with obvious nonlinear structure, which illustrates that TGBLDA inherits the nonlinear characteristics of the GbFA.

(3) Although TLLDA fuses linear discriminant information with the weight way of the local feature information, having the ability for capturing local structure features, the performance of TLLDA still significantly lower than that of TGBLDA , especially on YaleB. This shows that TGBLDA has the more capability for preserving the local structure feature information.

5. Conclusions

On the basis of Graph-based Fisher Analysis (GbFA), th

paper proposed a face recognition Tensor Graph-based Linear Discriminant Analysis (TGbLDA). The algorithm does not need to convert a feature matrix into a vector, inheriting the characteristics of local preserving and power discriminant from GbFA. Compared with the existing tensor linear discriminant analysis algorithms, the algorithm has the more ability for preserving local structure and power discriminant performance. However, for TGbLDA which fuses graph-based local learning, how to choose the best local neighbor parameter k for the most performance is still a major problem in TGbLDA is the next step in the next work.

Acknowledgments

The work is support by NSF of Zhejiang province in China (LQ12F02007) and the reform project of the new century of Zhejiang province in China (YB2010092).

References

- [1] McLachlan. (2004). Discriminant Analysis and Statistical, *Pattern Recognition*, Wiley Interscience.
- [2] Fisher, R. A. (1936), The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, 7,179-188.
- [3] Shuiwang J., Jieping, Y. (2008). Generalized linear discriminant analysis: a unified frame-work and efficient model selection, *IEEE Transactions on Neural Networks*, 19 (10) 1768-1782.
- [4] Juwei L., K. N. P., A. N. V. (2005). Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition, *Pattern Recognition Letters*, 26 (2) 181-191.
- [5] Trevor H., Andreas, B., Robert, T. (1995). Penalized discriminant analysis, *The Annals of Statistics*, 23 (1) 73-102.
- [6] Howland, P., Park, H. (2004). Generalizing discriminant analysis using the generalized singular value decomposition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 (8) 995-1006.
- [7] Jieping, Y., Qi, L. (2005). A two-stage linear discriminant analysis via QR-decomposition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (6) 929-941.

- [8] Jieping Y., Tao X. (2006). Computational and theoretical analysis of null space and orthogonal linear discriminant analysis, *Jouranl of Machine Learning Research*, 7, 1183-1204.
- [9] Lifen C., Hongyuan, M., Mingtat, K., Jachen, L., Gwojong, Y. (2000). A new LDA-based face recognition system which can solve the small sample size problem, *Pattern Recognition*, 33 (10) 1713-1726.
- [10] Hua, Y., Jie Y. (2001). A direct LDA algorithm for high dimensional data-with application to face recognition, *Pattern Recognition*, 34 (l) 2067-2070.
- [11] Zhifeng, L., Dahua, L., Xiaoou, T. (2009). Nonparametric Discriminant Analysis for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31 (4) 755-761.
- [12] Masashi S. (2006). Local Fisher discriminant analysis for supervised dimensionality reduction, *In: Proc. of the 23th International Conference on Machine Learning*.
- [13] Hua, W., Chris, D., Heng, H. (2010). Multi-label Linear Discriminant Analysis, *In: Proc. Of the 11th European Conference on Computer Vision*, p. 126-139.
- [14] Fan, Z, Xu, Y, Zhang, D. (2011). Local Linear Discriminant Analysis Framework Using Sample Neighbors, *IEEE Transactions on Neural Networks*, 22 (7) 1119-1132.
- [15] Yan, C., Liya, F. (2012). A novel supervised dimensionality reduction algorithm: Graph-based Fisher analysis, *Pattern Recognition*, 45 (4) 1471-1481.
- [16] Xiaofei, H., Deng, C., Partha, N. (2005). Tensor subspace analysis, *In: Proc. of the 18th Neural Information Processing Systems (NIPS)*.
- [17] Zhao, Z., Chow, W. S. (2011). Tensor Locally Linear Discriminative Analysis, *IEEE Signal Processing Letters*, 18 (11) 643-646.

Author Biography



Jian-Jun Chen received master degree from Taiyuan University of Technology, associate professor. His research fields include machine learning and image processing. chengjianjun_2012@163.com