Regularized Orthogonal Local Fisher Discriminant Analysis

Shuhua Xu
Department of Maths
University of Shaoxing
900 City South Road, Shaoxing
P.R China
webqmm1974@163.com

Journal of Digital
Information Management

ABSTRACT*: Aiming at deficiencies of the ability for preserving local nonlinear structure of recently proposed Regularized Orthogonal Linear Discriminant Analysis (ROLDA) for dimensionality reduction, a kind of dimensionality reduction algorithm named Regularized Orthogonal Local Fisher Discriminant Analysis (ROLFDA) is proposed in the paper, which is originated from ROLDA. The algorithm introduce the idea of local structure preserving in Local Fisher Discriminant Analysis (LFDA) on the basic of ROLDA, following properties of ROLDA and strengthening the ability for capturing local structure information of data with nonlinear structures. Experiments on real face datasets demonstrate the effectiveness of our proposed algorithm.*

Categories and Subject Descriptors:
G.4: [**Mathematical analysis**]**:** Nonlinear equations; **I.1.2:** [**Symbolic and algebraic algorithms**]

General Terms:
Fisher Discriminant Analysis, Nonlinear equations

Keywords: Dimensionality Reduction, Regularized Orthogonal Linear Discriminant Analysis, Local Nonlinear Structure, Local Fisher Discriminant Analysis

Received: 18 October 2012, Revised 2 December 2012, Accepted 12 December 2012

1. Introduction

In applications of data mining, high-dimensional data lead to too much redundant feature information and increase the computational complexity of disposing. Hence dimensionality reduction is necessary. The purpose of dimensionality reduction is to find a low-dimensional representation of the high-dimensional data at minimum cost in the process of preserving certain features. The most well-know initial dimensionality reduction methods are Principal Component Analysis (PCA) [1] and Linear Discriminant Analysis (LDA) [2]. As supervised dimensionality reduction, LDA searches the projection axes on which the not-same-class points are far from each other while requiring the same-class points to be close to each other. Therefore, LDA encodes discriminating information in a linearly separable space and is successfully applied into face recognition. However, it is much expensive to attain plenty of class labeled samples in practical applications of face recognition, which cause that the total scatter matrix can be singular in LDA since the data points are from a very high-dimensional space and thus usually the number of the data samples is much smaller than the data dimension. This is known as the undersampled problem. To solve the undersampled problem in LDA application, extensions of LDA can be proposed in the literature. The two-stage LDA is proposed in [3-5], which is to apply an intermediate dimensionality reduction stage to reduce the data dimensionality. Although this method is simple, some important information is removed in the intermediate dimensionality reduction stage. The regularized LDA (RLDA) is proposed in [6-10], which is to consider regularization by adding the perturbation to the total scatter matrix. The methods based on pseudoinverse [11] are applied to avoid the singularity problem, including orthogonal LDA (OLDA) [12-13], null space LDA (NLDA) [14-16], uncorrelated LDA [17], QR/GSVD-based LDA [18-20]. OLDA obtains easily the optimal transformation matrix by only orthogonal transformations without computing any eigen-decomposition and matrix inverse. Moreover, OLDA is implemented by using several QR factorizations and is a fast one.

In above extends of LDA for the undersampled problem, RLDA is an efficient method with little computation. The major problem of the RLDA is to choose an appropriate regularization parameter. Usually these RLDA algorithms select the regularization parameter from a given parameter candidate set by using cross-validation for classification. But how to choose an appropriate candidate set is not clear. Therefore, up to now, there is no concrete mathematical theory available in selecting an appropriate regularization parameter in practical applications of the RLDA. For filling this gap, Ching *et al* proposed Regularized Orthogonal Linear Discriminant Analysis (ROLDA) [21]. ROLDA derives from the mathematical relationship between OLDA and RLDA and finds a mathematical criterion for selecting the regularization parameter from the relation. However, the performance of ROLDA tends to be degraded on non-linear data because that ROLDA is originated from LDA which works very well if the samples in each class follow Gaussian distributions with a linear covariance structure. To overcome this drawback, Local FDA (LFDA) [22-23] has been proposed, which combines the ideas of FDA and LPP to maximize between-class reparability while preserve within-class local structure. LFDA can provide more separate embedding than FDA.

Inspired by LFDA, a kind of Regularized Orthogonal Local Fisher Discriminant Analysis (ROLFDA) is proposed for dimensionality reduction. The algorithm introduces the idea of LFDA into ROLDA, preserving local structure information of samples and calculating the regular parameter automatically. Therefore ROLFDA can be applied in high-dimensional data with non-linear structure. Experiments on YaleB and AR demonstrate the proposed algorithm is valid.

The rest of the paper is organized as follows: Section 2 reviews LDA, LFDA and ROLDA. Our ROLFDA is introduced in Section 3. In Section 4, we compare ROLFDA with some related works. The experimental results and analyses are presented. Finally, we provide some concluding remarks and future work in Section 5.

2. Related works

2.1 Linear Discriminant Analysis (LDA)

Given samples $X = \{x_i \mid x_i \in R^d\}_{i=1}^{n}$, containing C classes. n denotes the number of samples. The goal of LDA is to seek a projection matrix T such that the between-class scatter is maximized and the within-class scatter is minimized in projected data. The objective function of LDA is as follows:

$$T = \arg\max_{T} \operatorname{tr}\left((T^T S^w T)^{-1} (T^T S^b T)\right) \quad (1)$$

Where S^b denotes the between-class scatter matrix and S^w denotes the within-class scatter matrix S^b and S^w are defined as follow:

$$S^b = \frac{1}{n} \sum (\bar{x}_{ij} - \bar{x})(\bar{x}_i - \bar{x})^T \quad (2)$$

$$S^w = \frac{1}{n} \sum_{i=1}^{C} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T \quad (3)$$

Where x_{ij} denotes the j^{th} sample in the i^{th} class. $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ denotes the average of all samples X, $\bar{x} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$ denotes the average of samples in the i^{th} class. $n_i (1 \le i \le C)$ denotes the number of samples in the i^{th} class.

Equation (1) may be further transformed into following generalized eigenvalue problem:

$$S^b t_m = \lambda_m S^w t_m,\ m = 1, ..., l\ (l < d) \quad (4)$$

Get the projecting matrix $T = [t_1, t_2, ..., t_l]$.

2.2 Local Fisher Discriminant Analysis (LFDA)

LFDA is a localized variant of Fisher discriminant analysis. LFDA takes local structure of the data into account so the multimodaldata can be embedded appropriately. LFDA redefine $S^{(lb)}$ and $S^{(lw)}$ as follows:

$$S^{(lb)} = \frac{1}{2} \sum_{i,j=1}^{n} W_{i,j}^{(lb)} (x_i - x_j)(x_i - x_j)^T \quad (5)$$

$$S^{(lw)} = \frac{1}{2} \sum_{i,j=1}^{n} W_{i,j}^{(lw)} (x_i - x_j)(x_i - x_j)^T \quad (6)$$

where

$$W_{i,j}^{(lb)} = \begin{cases} A_{i,j}(1/n - 1/n'_{y_i}) & \text{if } y_i = y_j \\ 1/n & \text{if } y_i \neq y_j \end{cases} \quad (7)$$

$$W_{i,j}^{(lw)} = \begin{cases} A_{i,j}(1/n'_{y_i}) & \text{if } y_i = y_j \\ 0 & \text{if } y_i \neq y_j \end{cases} \quad (8)$$

$$A_{i,j} = \exp\left(-\frac{\|x_i - x_j\|^2}{\delta_i \delta_j}\right) \quad (9)$$

$$\delta_i = \|x_i - x_i^k\| \quad (10)$$

Where x_i^k is the k^{th} nearest neighbor of x_i, $A_{i,j}$ denotes the affinity weight of x_i and x_j, the parameter represents the local scaling around x_i. The objective function of LFDA is defined as follows:

$$T = \arg\max_{T} \operatorname{tr}\left((T^T S^{(lw)} T)^{-1} (T^T S^{(lb)} T)\right) \quad (11)$$

2.3 Regularized Orthogonal Linear Discriminant Analysis (ROLDA)

On the base of RLDA [7], the objective function of ROLDA is as follows:

$$T = \arg\max_{T} \operatorname{tr}\left((T^T S^w T + \lambda I)^{-1} (T^T S^b T)\right) \quad s.t.\ T^T T = I \tag{12}$$

Where, the appropriate regularization parameter λ is selected according to [12].

3. Regularized Orthogonal Local Fisher Discriminant Analysis (ROLFDA)

3.1 Basic idea

(1) RLDA is an efficient algorithm for the undersampled problem. However, the main task of the regularized LDA is to choose an appropriate regularization parameter λ. If is λ large, information on the scatter matrix is lost. While if it is too small, the regularization may not be sufficiently effective. Usually existing RLDA methods select the regularization parameterfrom a candidate set of the regularization parameter given with the cross-validation method. It has been shown in [30] that the matrix computations involved in the regularized LDA can be simplified so that the cross-validation procedure can be performed efficiently. However, for RLDA methods, it is not clear how to choose an appropriate candidate set. ROLDA makes the analysis of the relationship between OLDA and RLDA .By means of this relationship, ROLDA find a mathematical criterion for selecting the regularization parameter in ROLDA. Consequently ROLDA develops a new regularized orthogonal linear discriminant analysis method, in which no candidate set of regularization parameter is needed.

(2) In many practical applications of data mining, intrinsical low dimensional data often lie in a very high dimensional space. It is important to explicitly take into account the structure of the manifold on which the data may possibly reside. LPP is to find the optimal linear approximations to the eigen functions of the Laplace Beltrami operator on the manifold, sharing many of the data representation properties of nonlinear dimensionality reduction such as Laplacian Eigen maps or Locally Linear Embedding. LPP keeps nearby data pairs in the original space close in the embedding space, by which nonlinear data can be embedded without losing its local structure. LFDA combines the idea of LPP and FDA, inheriting merits of them.

As above analyses, on the base of ROLDA, the idea of local preserving in LFDA is introduced. This will overcome the poor ability of ROLDA for capturing the local structure and inherit characteristics of them.

3.2 Objective function

On the base of Equation (12), the objective function of ROLFDA is defined as follows:

$$T = \arg\max_{T} \operatorname{tr}\left((\lambda (T^T S^{(lb)} T) + (T^T S^{(lw)} T))^{-1} (T^T S^{(lb)} T)\right) \quad s.t.\ T^T T = I \tag{13}$$

Where, according to [11], the regularization parameter λ is solved with following steps:

(1) Given samples $X = \{x_i \mid x_i \in R^d\}_{i=1}^n$, containing k class face. Firstly, R and Q_1 are gotten via QR-decomposition as follows:

$$X = Q_1 \begin{bmatrix} R \\ 0 \end{bmatrix} \tag{14}$$

Where $Q_1 \in R^{d \times n}$ is column orthogonal.

(2) $X_1 \in R^{n \times 1}$, $X_2 \in R^{n \times (k-1)}$ and $X_2 \in R^{n \times (n-k)}$ are computed as follows:

$$[X_1\ X_2\ X_3] = R \begin{pmatrix} \vartheta_1 & & \\ & \ddots & \\ & & \vartheta_k \end{pmatrix} \rho \begin{bmatrix} \vartheta & \\ & 1 \end{bmatrix}$$

Where ρ be the permutation matrix obtained by exchanging the $\left(\sum_{j=1}^{i-1} n_j + 1\right)$ th-column of I_n and ith column (for $i = 2, \ldots, k$), but otherwise leaving the order of the remaining columns unchanged. $\vartheta_i (i = 2, \ldots, k)$ and ϑ are calculated as follows:

$$\vartheta_i = I - \left(\begin{bmatrix} 1-\sqrt{n_i} \\ 1 \\ \vdots \\ 1 \end{bmatrix} \Big/ \sqrt{n_i - \sqrt{n_i}}\right)\left(\begin{bmatrix} 1-\sqrt{n_i} \\ 1 \\ \vdots \\ 1 \end{bmatrix} \Big/ \sqrt{n_i - \sqrt{n_i}}\right)^T \tag{16}$$

$$i = 1, \ldots, k$$

$$\vartheta = I - \left(\begin{bmatrix} \sqrt{n_1}-\sqrt{n} \\ \sqrt{n_2} \\ \vdots \\ \sqrt{n_k} \end{bmatrix} \Big/ \sqrt{n_i - \sqrt{nn_1}}\right)\left(\begin{bmatrix} \sqrt{n_1}-\sqrt{n} \\ \sqrt{n_2} \\ \vdots \\ \sqrt{n_k} \end{bmatrix} \Big/ \sqrt{n_i - \sqrt{nn_1}}\right)^T \tag{17}$$

(3) calculate

$$[X_1\ X_2] = Q_1 \begin{bmatrix} R_{1,1} & R_{2,2} \\ 0 & R_{2,2} \\ 0 & 0 \end{bmatrix} \tag{18}$$

Where $R_{1,1} \in R^{q \times (k-1)}$ and $R_{2,2} \in R^{(\gamma - q) \times (n-k)}$ are of full row rank.

(4) calculate λ

$$\sqrt{\lambda} = \frac{\sqrt{\varepsilon}}{\|R_{2,2}^{(+)}\|_2 \sqrt{(1 + \|R_{1,2} R_{2,2}^{(+)}\|_2 [\varepsilon + 2(1+\sqrt{2})(\sqrt{n-q} + \|R_{1,2} R_{2,2}^{(+)}\|_F)])}} \tag{19}$$

Where ε denotes error, $R_{2,2}^{(+)}$ denotes the pseudo inverse of $R_{2,2}$.

3.3 Algorithm steps

Input: samples $X = \{x_i \mid x_i \in R^d\}_{i=1}^n$ with C class, error $\varepsilon > 0$.

Output: the column orthogonal projection matrix $T \in R^{d \times l}$.

Steps:

(1) calculate $W^{(lb)}$ 和 $W^{(lw)}$ using Equation (7) and Equation (8)

(2) calculate Q_1 using Equation (14).

(3) calculate X_1, X_2 and X_3 using Equation (15).

(4) calculate $R_{1,1}, R_{1,2}$ and $R_{2,2}$ using Equation (18).

(5) calculate λ using Equation (19).

(6) transform Equation (12) into following generalized matrix problem:,

$$S^{(lb)} t_m = \beta_m (\lambda S^{(lb)} + S^{(lw)}) t_m, \; m = 1, ..., l$$

and calculate $T = [t_1, t_2, ..., t_l]$.

4. Experimental results and analysis

4.1 Experimental Datasets and Settings

Face datasets are typical high-dimensional data with nonlinear structures. In the experiment, YaleB and AR are selected and are describe as follows:

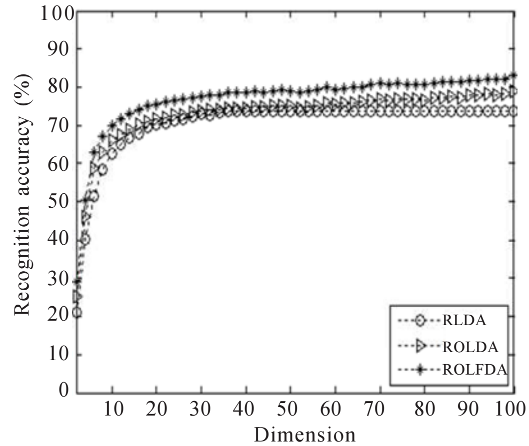
1) YaleB contains 2414 front-view face images of 38 individuals. For each individual, about 64 pictures were taken under various laboratory-controlled lighting conditions. In our experiments, The cropped images with the resolution of 32×32 are used. A group of samples in YaleB are shown in Figure 1.

2) AR database contains over 4000 face images of 126 individuals. For each individual, 26 pictures were taken in two sessions (separated by two weeks) and each section contains 13 images. These images include front view of faces with different expressions, illuminations and occlusions. In the experiment, them are resized to 66×48. A group of samples in AR are shown in Figure 2.

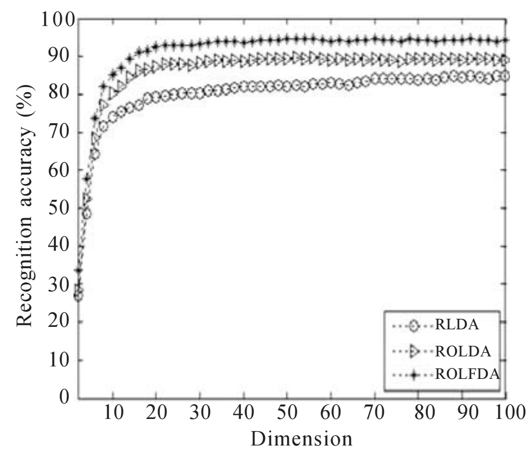
RLDA and ROLDA are compared with ROLFDA and the Nearest Neighbor Classifier is adopted in order to verify the performance of our algorithm. T images are selected randomly from a group face and remains for test samples. Moreover, the regularization parameter λ in RLDA is set to 0.1 and the neighbor parameter k in ROLFDA is set to 7.

4.2 Experimental Results

Experiments are made under different T and different reduced dimensions. All experiments are repeated twenty times and average recognition accuracy rates are gotten. Reduced dimensions are selected with the certain increment and corresponding average recognition accuracies are calculated. Experimental results are shown in Figure 3- Figure 4



(a) $T = 10$



(b) $T = 20$

Figure 3. Experimental results on YaleB with T

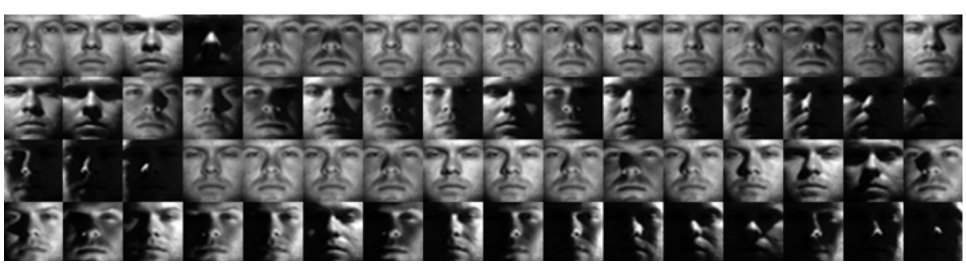
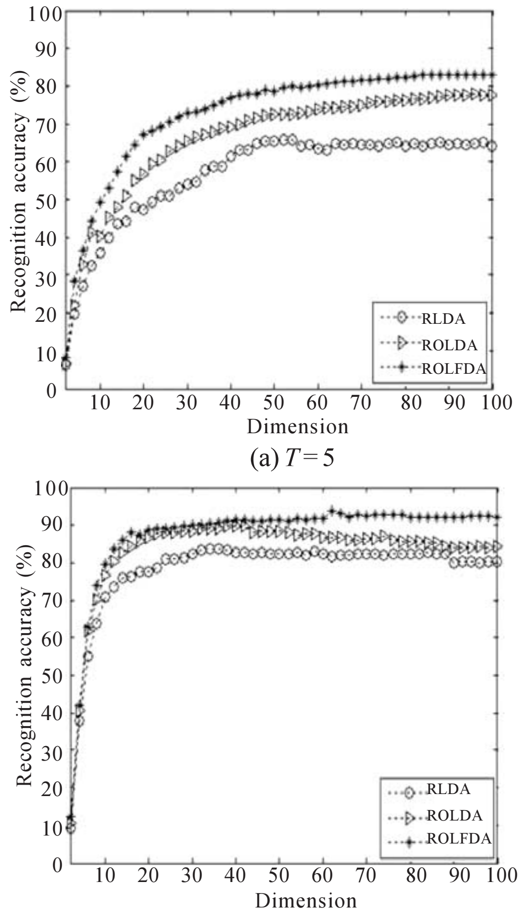


Figure 1. A group of samples in YaleB

Figure 2. A group of samples in AR



(a) $T = 5$

(a) $T = 10$

Figure 4. Experimental results on AR with T

From above Figure 3- Figure 4, in the term of the relation of recognition accuracy and reduced dimensions, conclusions can be drawn as follows:

(1) With increase in reduced dimension, the recognition accuracy rate of all algorithms promote rapidly firstly and go smoothly, which shows that these algorithms can get most recognition accuracy rates in low dimension .The reasons is that they is originated from LDA with the merit.

(2) ROLFDA is superior to RLDA and ROLDA in AR and YaleB with nonlinear structures, which is illustrated by that ROLFDA infuses efficiently the characteristics of preserving local structure in LFDA while RLDA and ROLDA fail to capture local structures.

Moreover, in the term of the relation of recognition accuracy and the number of training samples, some following conclusions are drawn:

1) As experiments on YaleB show, the most recognition accuracy rate of ROLFDA is respectively higher than that of RLDA and ROLDA when T is set to 10. The gap between them becomes larger when T is set to 20. This illuminate that more training samples is helpful for ROLFDA to capture the local structure feature.

2) The performance of ROLFDA is better under $T = 20$ than that under $T = 10$, which shows that more training samples make it for ROLFDA with the power ability for capturing local structure features to be more applied on AR with power external disturb, including illuminations and occlusions.

5. Conclusion

In the paper, Regularized Orthogonal Local Fisher Discriminant Analysis (ROLFDA) for dimensionality reduction is proposed. Aiming at the deficiency on the ability for capturing local structures information, ROLFDA introduce the idea of preserving local structures in Local Fisher Discriminant Analysis (LFDA) on the base of ROLDA, inheriting merits of ROLDA and promote the ability for local learning. Experimental results on YaleB and AR show the proposed algorithm fuse efficiently LFDA and ROLDA. However, for ROLFDA, the local structure preserving is based on the approximation of linearization instead of instinct geometrical structure. How to fuse other nonlinear algorithm in ROLFDA is our next work.

6. Acknowledgements

The research is supported by NSF of Zhejiang province (grant No. LQ12F02007, Y201122544).

References

[1] Fukunaga, K. (1990). Introduction to Statistical Pattern Recognition, 2nd ed., Academic Press, Boston, USA.

[2] Duda, R. O., Hart, P. E., Stork. D. G. (2000).Pattern Classification, second ed., John Wiley & Sons, New York.

[3] 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.

[4] Swets, D. L. ,Weng, J. J. (1996). Using discriminant eigenfeatures for image retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18 (8) 831–836.

[5] Ye, J. P., Li, Q. (2005). A two-stage linear discriminant analysis via QR-decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6) 929–941.

[6] Jerome, H. F. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, 84 (405) 165–175.

[7] Dai, D. Q., Yuen, P. C. (2003). Regularized discriminant analysis and its application to face recognition. *Pattern Recognition*, 36 (3) 845–847.

[8] Guo,Y. Q., Hastie, T. (2007). Regularized linear discriminant analysis and its application in microarray. *Biostatistics*, 8 (1) 86–100.

[9] Ye, P., Xiong,T., Li, Q., Janardan,R., Bi, B., Cherkassky V., Kambhamettu, C. (2006). Efficient model selection for regularized linear discriminant analysis. *In:* Proc. of the 15th ACM international conference on Information and knowledge management (CIKM'06), p. 532-539. ACM, November.

[10] Jieping, Y., Tie, W. (2006). Regularized discriminant analysis for high dimensional, low sample size data. *In:* Proc. of Proceedings of the 12th ACM SIGKDD International Conference on Knowledge discovery and Data Mining (KDD '06), p. 454–463. ACM, August.

[11] Golub, G. H., VanLoan, C. F. (1996). Matrix Computations, third ed.. The Johns Hopkins University Press, Baltimore, MD.

[12] Chu, D. L, Goh, S. T. (2010). A new and fast orthogonal linear discriminant analysis on undersampled problems. *SIAM Journal on Scientific Computing*, 32 (4) 2274–2297.

[13] Ye, J. P. (2005). Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems. *Journal of Machine Learning Research*, 6 (4) 483–502.

[14] Chen, L. F., Liao, H. Y., Ko, M. T., Lin, J. C., Yu, G. J. (2000). A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33 (10) 1713–1726.

[15] Chu, D., Thye, G. S. (2010). A new and fast implementation for null space based linear discriminant analysis. *Pattern Recognition*, 43 (4) 1373–1379.

[16] Rui, H., Qingshan, L., Hanqing, L., Songde, M. (2002). Solving the small sample size problem of LDA. *In: Proc. of the International Conference on Pattern Recognition* (ICPR'02), p. 29–32. IEEE Computer Society, April.

[17] Chu, D. L., Goh, S. T., Hung, Y. S. (2011). Characterization of all solutions for undersampled uncorrelated linear discriminant analysis problems. *SIAM Journal on Matrix Analysis and Applications*, 32 (2) 820–844.

[18] Peg, H., Moongu, J., Haesun, P. (2003). Structure preserving dimension reduction for clustered text data based on the generalized singular value decomposition. *SIAM Journal on Matrix Analysis and Applications* 25 (1) 165–179.

[19] Haesun, P., Drake, B. L., Sangmin, L., Park, C. H. (2007). Fast Linear Discriminant Analysis Using QR Decomposition and Regularization. *Technical Report GT-CSE-07-21*, Georgia Institute of Technology.

[20] Ye, J. P., Janardan, R., Park, C. H., Park, H. (2004). An optimization criterion for generalized discriminant analysis on undersampled problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 (8) 982–994.

[21] Chinga, W. K, Chub, D., Liaoc, L. Z, Wangb, X. Y (2012). Regularized Orthogonal Linear Discriminant Analysis. *Pattern Recognition*, 45 (7) 2719-2732.

[22] Masashi, S. (2006). Local Fisher discriminant analysis for supervised dimensionality reduction. *In:* Proc. of the 23rd international conference on Machine learning (ICML '06), p. 905-912. ACM, Apr.

[23] Masashi, S. (2007). Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *Journal of Machine Learning Research*, 8 (5) 1027-1061.

Author Biography

Shuhua XU was born in Jiangxi, China, in 1977. She received her M.S. degree from Guizhou University, Guizhou, China, in 2004. Now she is an instructor of the School of Mathematics Science and physics Science. Her research interests include optimization theory and application, machine learning theory and pattern recognition.