

Chapter 13

Local Reconstruction and Dissimilarity Preserving Semi-Supervised Dimensionality Reduction

Feng Li, Zhengqun Wang, Zhongxia Zhou and Wei Xue

Abstract In this paper, a semi-supervised dimensionality reduction algorithm for feature extraction, named LRDPSSDR, is proposed by combining local reconstruction with dissimilarity preserving. It focuses on local and global structure based on labeled and unlabeled samples in learning process. It sets the edge weights of adjacency graph by minimizing the local reconstruction error and preserves local geometric structure of samples. Besides, the dissimilarity between samples is represented by maximizing global scatter matrix so that the global manifold structure can be preserved well. Comprehensive comparison and extensive experiments demonstrate the effectiveness of LRDPSSDR.

Keywords Local reconstruction · Dissimilarity preserving · Semi-supervised dimensionality reduction · Face recognition

13.1 Introduction

Face recognition has become one of the most challenging problems in the application of pattern recognition. Face image is a high dimension vector, so numerous dimension reduction techniques have been proposed over the past few decades [1], in Principal component analysis (PCA) [2] and Linear discriminant analysis (LDA) [3] are widely used. Both PCA and LDA assume feature space lie on a linearly embedded manifold and aim at preserving global structure. However, many researches have shown that the face images possibly reside on a nonlinear submanifold [4]. When using PCA and LDA for dimensionality reduction, they will fail to discover the intrinsic dimension of image space.

F. Li (✉) · Z. Wang · Z. Zhou · W. Xue
Department of Automation, College of Information and Engineering,
Yangzhou University, Yangzhou 225127, China
e-mail: yzulifeng@126.com

By contrast, manifold learning considers the local information of samples, aiming to directly discover the globally nonlinear structure. The most important manifold learning algorithms include isometric feature mapping (Isomap) [5], locally linear embedding (LLE) [6], and Laplacian eigenmap (LE) [7]. Though these methods are appropriate for representation of nonlinear structure, they are implemented restrictedly on training samples and can not show explicit maps on new testing samples in recognition. Therefore, locality preserving projection (LPP) [8] is proposed, but it only can focus on the local information of training samples. To remedy this deficiency, unsupervised discriminant analysis (UDP) [9] is introduced, which can consider the local structure as well as global structure of samples. However, it only uses unlabeled data. Therefore, semi-supervised methods to deal with insufficient labeled data could be learned. It can be directly applied in the whole input space, while the out-of-sample problem can be effectively solved.

This paper simultaneously investigates two issues. First, How to extract the effective discriminant feature by using labeled samples? Second, How to minimize local scatter matrix and maximize global scatter matrix simultaneity, to preserve the local and global structure information?

The rest of this paper is organized as follows. In Sect. 13.2, we give the details of LRDPSDR. The experimental results based on two commonly used face databases demonstrate the effectiveness and robustness of proposed method in Sect. 13.3. Finally, the conclusions are summarized in Sect. 13.4.

13.2 Locally Reconstruction and Dissimilarity Preserving Semi-Supervised Dimensionality Reduction

13.2.1 Local Reconstruction Error

Suppose $\mathbf{X} = [x_1, x_2, \dots, x_l, x_{l+1}, \dots, x_{l+u}]$ be a set of training samples that include l labeled samples and u unlabeled samples, belonging to c classes. For each sample, we find its k nearest neighbors $N_k(x_i)$ from \mathbf{X} , where $N_k(x_i)$ is the index set of the k nearest neighbors of x_i . To preserve local structure, design following objective function:

$$J(a) = \sum_{i=1}^{l+u} \left\| x_i - \sum_{j: x_j \in N_k(x_i)} C_{ij} x_j \right\|^2 \quad (13.1)$$

where $\|\bullet\|$ denotes the Euclidean norm, with two constraints:

- (1) $\sum_j C_{ij} = 1, i = 1, 2, \dots, l + u$
- (2) $C_{ij} = 0$, if x_j does not belong to the set of k nearest neighbors of x_i .

Given samples in low-dimensional linear embedding space as follows:

$$\begin{aligned}
\sum_{i=1}^{l+u} \left\| y_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} y_j \right\|^2 &= \sum_{i=1}^{l+u} \left(\mathbf{w}^T x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} \mathbf{w}^T x_j \right)^T \left(\mathbf{w}^T x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} \mathbf{w}^T x_j \right) \\
&= \sum_{i=1}^{l+u} \left(\mathbf{w}^T \left(x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} x_j \right) \right)^T \left(\mathbf{w}^T \left(x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} x_j \right) \right) \\
&= \text{tr} \left\{ \mathbf{W}^T \left[\sum_{i=1}^{l+u} \left(x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} x_j \right) \left(x_i - \sum_{j:x_j \in N_k(x_i)} C_{ij} x_j \right)^T \right] \mathbf{W} \right\} \\
&= \text{tr}(\mathbf{W}^T \mathbf{S}_L \mathbf{W})
\end{aligned} \tag{13.2}$$

where $\text{tr}(\bullet)$ is the notation of trace. Local scatter matrix is defined as follows:

$$\mathbf{S}_L = \mathbf{X} \mathbf{M} \mathbf{X}^T \tag{13.3}$$

where $\mathbf{M} = (\mathbf{I} - \mathbf{C})^T (\mathbf{I} - \mathbf{C})$, \mathbf{I} is an $(l+u) \times (l+u)$ identity matrix, and matrix \mathbf{C} is calculated by LLE [6] algorithm.

With above two constraints, the local reconstruction weight is invariant to rotation, rescaling, and translation. So, it can preserve local structure of samples.

13.2.2 Dissimilarity Preserving

Let $\mathbf{X} = [x_1, x_2, \dots, x_l, x_{l+1}, \dots, x_{l+u}]$ be a matrix and each column is a training samples, belonging to c classes. Represent any x_i in term of its projection sequence $G(x_i)$, coefficient vector $W(x_i, G(x_i))$, and residue $R(x_i, G(x_i))$. Suppose $G(x_i) = [g_{i1}, g_{i2}, \dots, g_{id}]$ are the eigenvectors corresponding to the first d largest eigenvalues in PCA. Thus, each image can be represented as a linear combination:

$$\hat{x}_i = w_{i1} g_{i1} + w_{i2} g_{i2} + \dots + w_{id} g_{id} \tag{13.4}$$

where \hat{x}_i denotes the approximation of x_i , and the residue $R(x_i, G(x_i))$ is given by

$$R(x_i, G(x_i)) = x_i - (w_{i1} g_{i1} + w_{i2} g_{i2} + \dots + w_{id} g_{id}) = x_i - \sum_{j=1}^d w_{ij} g_{ij} \tag{13.5}$$

When x_1 is projected on projection sequence $G(x_2)$ of x_2 , noting the corresponding coefficient vector $W(x_1, G(x_2))$ and residue $R(x_1, G(x_2))$. Based on above factor, we can design the dissimilarity measure as follows:

$$\varphi(x_1, x_2) = \sqrt{\xi D_R(x_1, x_2) + (1 - \xi) D_W(x_1, x_2)} \tag{13.6}$$

where $\zeta \in [0, 1]$ indicates the relative importance of the residue and the corresponding coefficients, when both are projected onto the projection sequence $G(x_2)$ of x_2 , we have

$$D_R(x_1, x_2) = \|R(x_1, G(x_2)) - R(x_2, G(x_2))\|^2 \quad (13.7)$$

where $D_R(x_1, x_2)$ is the difference between the residues of x_1 and x_2 , and $D_W(x_1, x_2)$ compares their corresponding coefficients, which is represented as:

$$D_W(x_1, x_2) = \|W(x_1, G(x_2)) - W(x_2, G(x_2))\|^2 \quad (13.8)$$

Define the objective function of dissimilarity preserving as follows:

$$\sum_{i=1}^{l+u} \sum_{j=1}^{l+u} (y_i - y_j)^2 \mathbf{H}_{ij} \quad (13.9)$$

where \mathbf{H}_{ij} is the dissimilarity weight matrix, and $\mathbf{H}_{ij} = \varphi(x_i, x_j)$.

$$\begin{aligned} \frac{1}{2} \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} (y_i - y_j)^2 \mathbf{H}_{ij} &= \frac{1}{2} \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} (\mathbf{W}^T x_i - \mathbf{W}^T x_j)^2 \mathbf{H}_{ij} = \mathbf{W}^T \frac{1}{2} \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} (x_i - x_j)(x_i - x_j)^T \mathbf{H}_{ij} \mathbf{W} \\ &= \mathbf{W}^T \frac{1}{2} \left(\sum_{i=1}^{l+u} \sum_{j=1}^{l+u} x_i x_i^T \mathbf{H}_{ij} + \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} x_j x_j^T \mathbf{H}_{ij} - 2 \sum_{i=1}^{l+u} \sum_{j=1}^{l+u} x_i x_j^T \mathbf{H}_{ij} \right) \mathbf{W} \\ &= \mathbf{W}^T (\mathbf{X} \mathbf{D} \mathbf{X}^T - \mathbf{X} \mathbf{H} \mathbf{X}^T) \mathbf{W} = \mathbf{W}^T \mathbf{S}_N \mathbf{W} \end{aligned} \quad (13.10)$$

Further, the global scatter matrix is defined as follows:

$$\mathbf{S}_N = \mathbf{X} \mathbf{L} \mathbf{X}^T \quad (13.11)$$

where $\sum \mathbf{H}_{ij}$ is a diagonal matrix, $\mathbf{L} = \mathbf{D} - \mathbf{H}$ is a Laplacian matrix.

As a result, by maximizing the global scatter matrix can make the nearby samples of the same class become as compact as possible and simultaneously the nearby samples belonging to different classes become as far as possible.

13.2.3 The Algorithm of LRDPSDR

According to above detail analysis, we have two scatter matrices based on labeled data and unlabeled data. Associated them with Fish criterion, a semi-supervised learning method will be derived. Design the objective function as follows:

$$\mathbf{W}_{opt} = \arg \max_{\mathbf{W}} \frac{\mathbf{W}^T (\mathbf{S}_b + \beta \mathbf{S}_N) \mathbf{W}}{\mathbf{W}^T (\mathbf{S}_t + \alpha \mathbf{S}_L) \mathbf{W}} = \arg \max_{\mathbf{W}} \frac{\mathbf{W}^T (\mathbf{S}_b + \beta \mathbf{X} \mathbf{L} \mathbf{X}^T) \mathbf{W}}{\mathbf{W}^T (\mathbf{S}_t + \alpha \mathbf{X} \mathbf{M} \mathbf{X}^T) \mathbf{W}} \quad (13.12)$$

where $\alpha, \beta > 0$ are the regularization parameters. The optimization problem solution of formula (13.12) could obtain by following eigen-equation:

$$(\mathbf{S}_b + \beta \mathbf{X} \mathbf{L} \mathbf{X}^T) \mathbf{W} = \lambda (\mathbf{S}_t + \alpha \mathbf{X} \mathbf{M} \mathbf{X}^T) \mathbf{W} \quad (13.13)$$

Based on above discussion, the proposed algorithm LRDPSDDR is briefly stated as below:

- Step 1: For the given training data set, use PCA for dimensionality reduction.
- Step 2: Calculate within-class scatter matrix \mathbf{S}_b and total scatter matrix \mathbf{S}_t for labeled data.
- Step 3: Construct local scatter matrix \mathbf{S}_L using formula (13.3) and global scatter matrix \mathbf{S}_N using formula (13.11).
- Step 4: The optimal transformation matrix \mathbf{W} is formed by the d eigenvectors corresponding to the first d largest non-zero eigen-value of formula (13.13).
- Step 5: Project the training data set onto the optimal projection vectors obtained in Step 4, and then use the nearest neighbor classifiers for classification.

13.3 Experiments

In this section, we investigate the performance of our algorithm for face recognition and compare it with PCA, LDA, LPP, and UDP. The KNN classifier is used. The regularization parameters are set as $\alpha = \beta = 0.1$, $\zeta = 0.5$.

13.3.1 Experiment on the Yale Dataset

The database consists of 165 face images of 15 individuals. These images are taken under different lighting condition and different facial expression. All images are gray scale and normalized to a resolution of 100×80 pixels. In the experiment, we select the first five images from each individual to form the training samples and the remaining six images as testing samples. Some typical images are shown in Fig. 13.1.

Figure 13.2 show that the recognition rate of LRDPSDDR method has significantly improvement compared to other four algorithms with the increase in the number of projection axis dimension. The projection axis dimension from 20 to 75 stages, the recognition rate tends stable. It can achieve a maximum value when projection axis dimension reach 55.

Table 13.1 gives the maximum recognition rate of five algorithms. It is not difficult to see that our method is the best according to Table 13.1 and Fig. 13.2.



Fig. 13.1 Sample images for one individual in the YALE database

Fig. 13.2 Comparison of recognition rate on YALE database

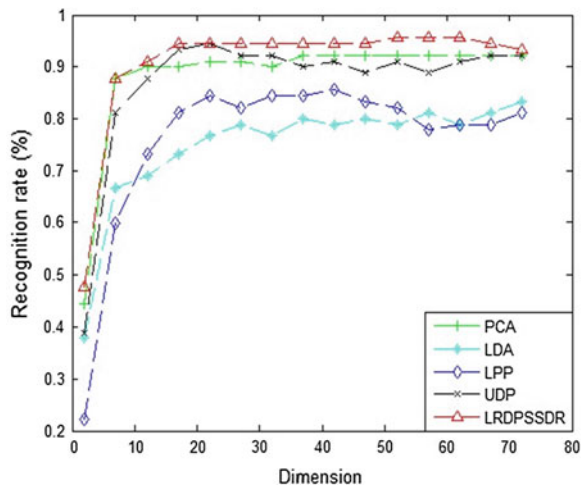


Table 13.1 Comparison on top accuracy

Algorithms	PCA	LDA	LPP	UDP	LRDPSDR
Recognition rate (%)	92.22	83.33	85.86	92.22	95.56
Dimension	35	75	40	70	55

Table 13.2 Comparison on top accuracy

Algorithms	PCA	LDA	LPP	UDP	LRDPSDR
Recognition rate (%)	62.98	58.81	61.19	64.52	70.71
Dimension	110	140	160	160	140



Fig. 13.3 Part of the sample images for one individual in the AR database

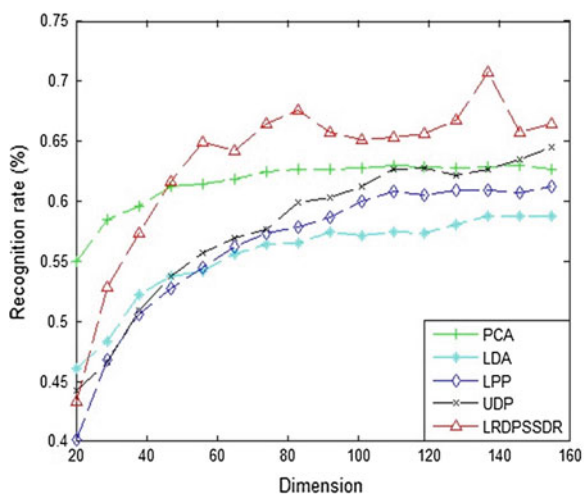
13.3.2 Experiment on the AR Dataset

The database consists of 3120 face images 120 individuals. These face images are captured under varying facial expressions, lighting conditions and occlusions. The size of every face image is gray scale and normalized to a resolution of 50×40 pixels. In the experiment, we select the first seven images from each individual to form the training samples and next seven images (most are covered) as testing samples. Some typical images are shown in Fig. 13.3.

From Fig. 13.4, we can see first that the proposed LRDPSSDR method outperforms PCA, LDA, LPP and UDP, and second that our method is more robust in different lighting conditions and various facial expressions. With the projection axis dimension increase, the recognition rate raises from 30 to 140 stages. It can achieve a maximum value when projection axis dimension is 140.

Table 13.2 shows the maximum recognition rate of five algorithms. It is obvious that our method is better than other methods, so the effectiveness and robustness of LRDPSSDR is further verified.

Fig. 13.4 Comparison of recognition rate on AR database



13.4 Conclusion

In this paper, we present a semi-supervised learning algorithm LRDPSDDR for dimension reduction, which can make use of both labeled and unlabeled data. The algorithm is realized based on both local reconstruction error and dissimilarity preserving, which not only preserves the intraclass compactness and the interclass separability, but also describes local and global structure of samples.

Acknowledgments We wish to thank the National Science Foundation of China under Grant No. 61175111, and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant No. 10KJB510027 for supporting this work.

References

1. Cevikalp H, Neamtu M, Wilkes M, Barkana A (2005) Discriminative common vectors for face recognition. *IEEE Trans Pattern Anal Mach Intell* 27(1):4–13
2. Turk M, Pentland A (1991) Eigenface for recognition. *J Cogn Neurosci* 3(3):72–86
3. Zheng WS, Lai JH, Yuen PC (2009) Perturbation LDA: learning the difference between the class empirical mean and its expectation. *Pattern Recogn* 42(5):764–779
4. Chen H-T, Chang H-W, Liu T-L (2005) Local discriminant embedding and its variants. In: *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, vol 2, pp 846–853
5. Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–2323
6. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290:2323–2326
7. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput* 15(6):1373–1396
8. He XF, Niyogi, P (2003) Locality preserving projections. In: *Proceedings of the 16th annual neural information processing systems conference*, Vancouver, pp 153–160
9. Yang J, Zhang D (2007) Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics. *IEEE Trans Pattern Anal Intell* 29(4):650–664