Locality Preserving Discriminant Projections

Jie Gui[1,2,*], Chao Wang[1,2], and Ling Zhu[1,2]

[1] Intelligent Computing Lab, Institute of Intelligent Machine, Chinese Academy of Science, P.O. Box 1130, Hefei, Anhui, 230031, China
[2] Departments of Automation, University of Science and Technology of China, Hefei, Anhui, 230027, China
guijiejie@gmail.com

Abstract. A new manifold learning algorithm called locality preserving discriminant projections (LPDP) is proposed by adding between-class scatter matrix and within-class scatter matrix into locality preserving projections (LPP). LPDP can preserve locality and utilize label information in the projection. It is shown that the LPDP can successfully find the subspace which has better discrimination between different pattern classes. The subspace obtained by LPDP has more discriminant power than LPP, and is more suitable for recognition tasks. The proposed method was applied to USPS handwriting database and compared with LPP. Experimental results show the effectiveness of the proposed algorithm.

Keywords: Manifold learning, locality preserving discriminant projections, locality preserving projections.

1 Introduction

In the past few years, the computer vision and pattern recognition community has witnessed the rapid growth of a new kind of feature extraction method, namely, manifold learning. Locally linear embedding (LLE) [1-3], isometric feature mapping (Isomap) [4, 5], and Laplacian eigenmap [6, 7] are the three most famous manifold learning algorithms. But these approaches have limitations and are not suitable for pattern recognition because they cannot give an explicit subspace mapping and thus cannot deal with the out-of-sample problem directly.

Many linear approaches have been proposed for dimensionality reduction, such as principal component analysis (PCA) [8, 9], multidimensional scaling (MDS) [10] and so on. All of these methods are easy to implement. At the same time, their optimizations are well understood and efficient. Because of these advantages, they have been widely used in visualization and classification. Unfortunately, they have a common inherent limitation that they all deemphasize discriminant information, which is important in recognition problem.

He et al. [11] proposed a new linear dimensionality reduction method named locality preserving projections (LPP) and applied it in pattern recognition tasks successfully [12]. LPP is originally derived by the linear approximation of the Laplace Beltrami

* Corresponding author.

D.-S. Huang et al. (Eds.): ICIC 2009, LNAI 5755, pp. 566–572, 2009.
© Springer-Verlag Berlin Heidelberg 2009

operator on a compact Riemannian manifold. However, LPP suffers from a limitation that it does not encode discriminant information while discriminant information is important for recognition task.

By introducing between-class scatter and within-class scatter into the objective function of LPP, we propose a locality preserving discriminant projections (LPDP) algorithm which not only considers the local information but also emphasizes the discriminant information. Experimental results show it is more suitable for recognition tasks with labeled information.

The remainder of this paper is organized as follows: Section 2 reviews manifold learning algorithms; Experiments and results are presented in Section 3; Conclusions are given in Section 4.

2 Mainfold Learning

The algorithm of linear dimensionality reduction finds a transformation matrix V that maps these N samples to a set of weight vectors $Y = [y_1, y_2, \cdots y_N]$, where Y is $d \times N$ matrix($d \ll D$). A reasonable criterion for choosing a 'good' map is to minimize an objective function [6].

2.1 Locality Preserving Projections (LPP)

The objective function of LPP [11, 12] is defined as:

$$\min \sum_{ij} \left\| Y_i - Y_j \right\|^2 W_{ij} , \tag{1}$$

where $Y_i = V^T X_i$ and the matrix $W = \left(W_{ij} \right)$ is a similarity matrix. The weight W_{ij} incurs a heavy penalty when neighboring points X_i and X_j are mapped far apart. Therefore, minimizing the objective function is an attempt to ensure that if X_i and X_j are "close" then Y_i and Y_j are close as well. A possible way of defining W is as follows:

$W_{ij} = 1$, if X_i is among k nearest neighbors of X_j or X_j is among k nearest neighbors of X_i, otherwise, $W_{ij} = 0$. k is the number of neighbors. The justification for this choice of weights can be traced back to [6].

Suppose V is a transformation matrix, that is, $Y = V^T X$. By simple algebra formulation, the objective function can be reduced to:

$$\begin{aligned} &\frac{1}{2} \sum_{ij} \left\| Y_i - Y_j \right\|^2 W_{ij} = \frac{1}{2} trace\Big(\sum_{ij} (V^T x_i - V^T x_j)(V^T x_i - V^T x_j)^T W_{ij}\Big) \\ &= trace\Big(\sum_{ij} V^T x_i D_{ii} x_i^T V - \sum_{ij} V^T x_i W_{ij} x_j^T V\Big) = trace(V^T X (D - W) X^T V) \end{aligned} , \tag{2}$$

where $X = [x_1, x_2, \cdots x_N]$, D is a diagonal matrix and D_{ii} is column (or row) sum of W, $D_{ii} = \sum_j W_{ij}$, $L = D - W$ is the Laplacian matrix.

In addition, a constraint is proposed as $V^T XDX^T V = I$.
Finally, the minimization problem reduces to finding:

$$\begin{gathered} \arg\ \min \operatorname{trace}(V^T XLX^T V) \\ V^T XDX^T V = I \end{gathered} \quad . \tag{3}$$

The transformation matrix that minimizes the objective functions is given by the minimum eigenvalues solution to the generalized eigenvalues problem:

$$XLX^T v = \lambda XDX^T v \, . \tag{4}$$

It is easy to show that the matrices XLX^T and XDX^T are symmetric and positive semidefinite. The vectors υ_i that minimize the objective function are given by minimum eigenvalues solutions to the generalized eigenvalues problem. Let the column vectors $\upsilon_0, \upsilon_1, \cdots \upsilon_{d-1}$ be the solutions of Eq. (4), ordered according to their eigenvalues, $\lambda_0, \lambda_1, \cdots \lambda_{d-1}$. Thus, the embedding is as follows

$$x_i \rightarrow y_i = V^T x_i, V = [\upsilon_0, \upsilon_1, \cdots \upsilon_{d-1}] \, . \tag{5}$$

Where y_i is a d-dimensional vector, and V is a $D \times d$ matrix.

2.2 Maximizing Margin Criterion (MMC)

Recently, MMC [13, 14] was proposed to determine the optimized subspace, which can maximize the average margin between classes after dimensionality reduction and can also successfully conquer the SSS (small sample size) problem. The objective function of MMC is written as:

$$\text{J=max}\left\{ \sum_{ij} p_i p_j \left(d\left(m_i, m_j\right) - s\left(m_i\right) - s\left(m_j\right) \right) \right\} . \tag{6}$$

Where p_i and p_j are the prior probability of class i and class j, m_i and m_j are the centroids of class i and class j. $d\left(m_i, m_j\right)$, $s\left(m_i\right)$ and $s\left(m_j\right)$ have the following definitions:

$$d\left(m_i, m_j\right) = m_i - m_j \ , \tag{7}$$

$$s\left(m_i\right) = tr\left(S_i\right) \ , \tag{8}$$

$$s\left(m_j\right) = tr\left(S_j\right) . \tag{9}$$

Thus the optimized function can be derived as follows:

$$J = max\, tr\left(S_b - S_w\right) . \tag{10}$$

The matrix S_b is called between-class scatter matrix and S_w is called within-class scatter matrix.

2.3 Locality Preserving Discriminant Projections (LPDP)

If the linear transformation obtained by LPP can satisfy Eq. (3) simultaneously, the discriminability of the data will be improved greatly. Thus the problem can be represented as the following multi-objective optimized problem:

$$\begin{cases} \min\ \text{trace}\left(V^T XLX^T V\right) \\ \max\ \text{trace}\left(V^T \left(S_b - S_w\right) V\right) \end{cases} \tag{11}$$
$$s.t.\ \ V^T XDX^T V = I .$$

It can be changed into the following constrained problem:

$$\begin{aligned} &\min\ \text{trace}\left(V^T \left(XLX^T - \alpha\left(S_b - S_w\right)\right) V\right) \\ &s.t.\ \ V^T XDX^T V = I . \end{aligned} \tag{12}$$

The parameter α can adjust the contributions of maximizing margin criterion and the local structure according to different data sets. The transformation matrix V that minimizes the objective function can be obtained by solving the generalized eigenvalue problem:

$$\left(XLX^T - \alpha\left(S_b - S_w\right)\right)V = \lambda XDX^T V . \tag{13}$$

3 Experimental Results

The proposed algorithm is here applied to USPS handwriting database. The USPS handwriting digital data [6] include 10 classes from “0” to “9”. Each class has 1100 examples. In our experiment, we select a subset from the original database. We cropped each image to be size of 16×16. There are 100 images for each class in the subset and the total number is 1000.

For each digit, l= (30, 40, 50, 60, 70) data are randomly selected for training and the rest are used for testing. In the PCA phase of LPP and LPDP, we kept 100 percent energy. The K-nearest neighborhood parameter k in LPP and LPDP was taken to be 5. The parameter α was taken to be 1. Finally, a nearest neighbor classifier with

Euclidean distance is employed to classify in the projected feature space. For each given *l*, we repeated this process for 20 times. Table 1 shows the best result obtained in the optimal subspace and the corresponding dimensionality. The best mean recognition rates for 20 times, standard deviation, and the corresponding dimensionality are also shown in Table 1. Fig. 1 displays the mean recognition rate curves versus the subspace dimensions when *l*= (30, 40, 50, 60).

Table 1. The maximal recognition rates, maximal average recognition rates, the corresponding standard deviations (percent) with the reduced dimensions for LPP, LPDP on USPS handwriting database (a nearest neighbor classifier with Euclidean distance)

Method	Accuracy (%)	30 Train	40 Train	50 Train	60 Train	70 Train
LPP	max(dimension)	43.86(20)	64.83(22)	74(12)	78.5(16)	84(24)
	average(dimension)	39.01(20)	58.58(20)	68.41(16)	74.49(20)	78.38(14)
	std	2.75	2.17	1.6	2.41	2.16
LPDP	max(dimension)	49.86(10)	70.67(12)	78.8(12)	83.25(14)	87.33(14)
	average(dimension)	46.44(10)	65.29(12)	73.5(12)	79.51(14)	82.32(14)
	std	2.54	2.54	1.64	2.04	2.01

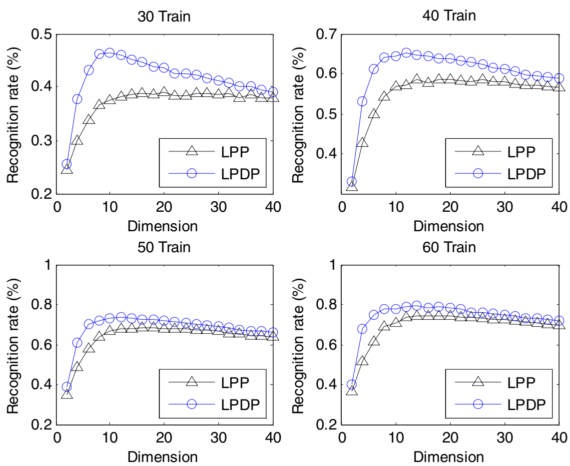

Fig. 1. The performance comparison of average recognition rates using LPP and LPDP by varying the dimensions for 30 Train, 40 Train, 50 Train, 60 Train (a nearest neighbor classifier with Euclidean distance)

If we employ a nearest neighbor classifier with cosine distance, the results will be better than a nearest neighbor classifier with Euclidean distance. This is consistent with the observation in [15]. We did not show them here due to limited space. If we employ a nearest neighbor classifier with cosine distance and normalize each image to be a unit vector, the results will be the best of all. The best results are reported in Table 2 and Figure 2.

Table 2. The maximal recognition rates, maximal average recognition rates, the corresponding standard deviations (percent) with the reduced dimensions for LPP, LPDP on USPS handwriting database (a nearest neighbor classifier with cosine distance)

Method	Accuracy(%)	30 Train	40 Train	50 Train	60 Train	70 Train
LPP	max(dim)	53.14(32)	73.00(40)	81.00(22)	84.50(30)	88.67(38)
	avg(dim)	48.01(40)	66.84(32)	75.65(36)	79.97(36)	84.07(30)
	std	2.27	1.73	1.85	2.49	1.93
LPDP	max(dim)	57.14(24)	75.50(22)	82.80(34)	88.00(28)	91.00(40)
	avg(dim)	51.24(36)	71.33(30)	79.46(26)	83.41(34)	86.98(26)
	std	2.96	1.52	1.42	2.24	1.76

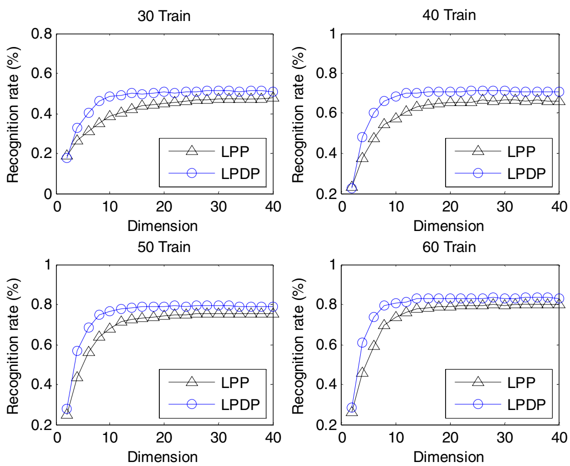

Fig. 2. The performance comparison of average recognition rates using LPP and LPDP by varying the dimensions for 30 Train, 40 Train, 50 Train, 60 Train (a nearest neighbor classifier with cosine distance)

Both Table 1,2 and Fig. 1,2 show that our LPDP algorithm performed better than LPP for all cases.

4 Conclusion

A novel method called LPDP is proposed. The new method adds the maximum margin criterion into the target function of LPP. The experiments on USPS handwriting database show that LPDP can improve the performance of LPP remarkably.

Acknowledgments

This work was supported by the grant from the National Basic Research Program of China (973 Program), No. 2007CB311002.

References

1. Roweis, S.T., Saul, L.K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 290(5500), 2323–2326 (2000)
2. Saul, L.K., Roweis, S.T.: Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds. Journal of Machine Learning Research 4(2), 119–155 (2004)
3. Li, B., Zheng, C.H., Huang, D.S.: Locally Linear Discriminant Embedding: An Efficient Method for Face Recognition. Pattern Recognition 41(12), 3813–3821 (2008)
4. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290(5500), 2319–2324 (2000)
5. Li, B., Huang, D.S., Wang, C.: Improving the Robustness of ISOMAP by De-noising. In: Proceedings of 2008 IEEE World Congress on Computational Intelligence (WCCI 2008), Hong Kong, pp. 266–270 (2008)
6. Belkin, M., Niyogi, P.: Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In: Advances in Neural Information Processing Systems 14, vols. 1, 2, pp. 585–591 (2002)
7. Belkin, M., Niyogi, P.: Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural Computation 15(6), 1373–1396 (2003)
8. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
9. Martinez, A.M., Kak, A.C.: PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), 228–233 (2001)
10. Williams, C.K.I.: On a Connection between Kernel PCA and Metric Multidimensional Scaling. Machine Learning 46(1-3), 11–19 (2002)
11. He, X.F., Niyogi, P.: Locality Preserving Projections. Advances in Neural Information Processing Systems 16 16, 153–160 (2004)
12. He, X.F., Yan, S.C., Hu, Y.X., et al.: Face Recognition Using Laplacianfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(3), 328–340 (2005)
13. Liu, J., Cheri, S.C., Tan, X.Y., et al.: Comments on Efficient and Robust Feature Extraction by Maximum Margin Criterion. IEEE Transactions on Neural Networks 18(6), 1862–1864 (2007)
14. Li, H.F., Jiang, T., Zhang, K.S.: Efficient and Robust Feature Extraction by Maximum Margin Criterion. IEEE Transactions on Neural Networks 17(1), 157–165 (2006)
15. Yang, J., Zhang, D., Yang, J.Y., et al.: Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Applications to Face and Palm Biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(4), 650–664 (2007)