A New Solution Scheme of Unsupervised Locality Preserving Projection Method for the SSS Problem

Yong Xu[1] and David Zhang[2]

[1] Harbin Institute of Technology, Shenzhen Graduate School
518055 Shenzhen, China
laterfall2@yahoo.com.cn
[2] The Biometrics Research Centre, The Hong Kong Polytechnic University
Kowloon, Hong Kong, China
csdzhang@comp.polyu.hk

Abstract. When locality preserving projection (LPP) method was originally proposed, it takes as the LPP solution the minimum eigenvalue solution of an eigenequation. After that, LPP has been used for image recognition problems such as face recognition. However, almost no researcher realizes that LPP usually encounters several difficulties when applied to the image recognition problem. For example, since image recognition problems are usually small sample size (SSS) problems, the corresponding eigenequation cannot be directly solved. In addition, it seems that even if one can obtain the solution of the eigenequation by using the numerical analysis approach, the obtained conventional LPP solution might produce the 'presentation confusion' problem for samples from different classes, which is disadvantageous for the classification to procedure a high accuracy. In this paper we first thoroughly investigate the characteristics and drawbacks of the conventional LPP solution in the small sample size (SSS) problem in which the sample number is smaller than the data dimension. In order to overcome these drawbacks, we propose a new LPP solution for the SSS problem, which has clear physical meaning and can be directly and easily worked out because it is generated from a non-singular eigenequation. Experimental results the proposed solution scheme can produce a much lower classification error rate than the conventional LPP solution.

Keywords: Locality preserving projection; Feature extraction; Face recognition.

1 Introduction

LPP is well-known as a linear graph embedding method. While LPP transforms different samples into their respective new representations using a same linear transform, it tries its best to preserve the local structure of the samples [1]--[12]. That it, after the LPP transformation, close samples in the original space is still close in the new space. LPP was firstly proposed as an unsupervised method that does not exploit the class-label information. So far a number of improvements of LPP such as kernel-based LPP [10], [13], two-dimensional LPP [14] and orthogonal LPP [15] have

N. da Vitora Lobo et al. (Eds.): SSPR&SPR 2008, LNCS 5342, pp. 775–781, 2008.

been developed. Recently, it has been demonstrated that LPP is theoretically related to other linear dimensionality reduction methods. In other words, these methods can be described as implementations of the linear graph embedding framework with different choices of the weight matrix and a related matrix [16].

Conventional LPP, which takes as the optimal solution of LPP the minimum eigenvalue solution of an eigenequation, will suffer from several difficulties in the SSS problem, in which the data dimension is much larger than the number of samples. The first difficulty is as follows: since the dimension of the sample is larger than the number of the samples, the eigenequation cannot be directly solved due to matrix singularity. An image recognition problem such as face recognition is a typical SSS problem. Note that image recognition covers a wide range of pattern recognition problems; therefore, the study to how to properly apply LPP to the SSS problem is very significant. But to our best knowledge, almost no researcher has worked for this and no satisfactory approach is proposed. It seems that existing LPP-based image recognition applications are usually interested in avoiding the SSS problem rather than finding a good solution to this problem. For example, most of face recognition applications of LPP firstly reduce the size of the face image and then implement the LPP algorithm based on the resized images. Theoretically, to make the LPP algorithm be workable, the dimension of the one-dimensional vector of the resized image should be smaller than the number of the training samples. Consequently, in order to make the dimension of the image not be larger than the sample number to avoid the SSS problem, the original image usually should be resized into a quite low size. Therefore, the process of reducing the size of the face image will case a large quantity of image information loss. This is disadvantageous for recognition to obtain a high accuracy.

The second difficulty of conventional LPP in the SSS problem is that the minimum eigenvalue solution seems not to be the genuine optimal solution for the purpose of locality preserving projection. The reason is as follows. According to the theory of eigenvalue, if a real symmetric matrix is not full rank, it has zero eigenvalues and the number of zero eigenvalues is the same as the result of its dimension subtracted by the rank. As a result, in the SSS problem, even if the eigenequation can be numerically solved by employing the numerical analysis approach, there will be a large number of zero eigenvalues. Conventional LPP will take as transforming axes the eigenvectors corresponding to the zero eigenvalues. Consequently, after conventional LPP transforms samples into new representations using these transforming axes, a sample statistically will have the same representation as its 'neighbors', as demonstrated in the context below. Indeed, in this case, even if the 'neighboring' samples are from different same classes, conventional LPP will still produce the same representation for them, which is harmful for the classification procedure.

In this paper, we propose a new solution scheme for LPP. The new solution scheme does not suffer from the same difficulties as conventional LPP. Moreover, the new solution scheme appears to be subject to the motivation of locality preserving projection and has clear justification. The remainder of the paper is organized as follows. In Section 2 we introduce conventional LPP and its difficulties in the SSS problem. In Section 3 we present our LPP solution scheme and show its advantages. In Section 4 we describe the experimental results. Finally we offer our conclusion in Section 5.

2 Algorithm and Analysis of Unsupervised LPP

LPP was proposed to transform samples into a new space by use a linear transform and to make close samples from the original space be still close in the new space. According to the objective function of LPP, its optimal transforming axis should be the minimum eigenvalue solution z of the following eigenequation:

$$(XDX^T)^{-1}XLX^T z = \lambda z \tag{1}$$

where $L = D - W$, X stands for the matrix consisting of n training samples $x_1 \quad x_2 \ldots x_n$ i.e. $X = [x_1 \quad x_2 \ldots x_n]$. D is defined as $D_{ii} = \sum_j w_{ij}$. Note that D is a dialog matrix and W is a symmetric matrix and the element w_{ij} of W is defined as follows: if x_j (or x_i) is one of k neighbors of x_i (or x_j), then $w_{ij} = \exp(-\|x_i - x_j\|^2/t)$; otherwise, $w_{ij} = 0$. Both L and D are positive semi-definitive matrices. For simplicity of presentation, hereafter we define that $D_1 = XDX^T$ and $L_1 = XLX^T$.

We call the eigenvector corresponding to the minimum eigenvalue of (1) conventional LPP solution and call the solving scheme conventional LPP. For real-world applications, conventional LPP usually firstly sorts the eigenvectors in increasing order of the corresponding eigenvalues and takes the first number of eigenvectors as transforming axes to implement the LPP transform. The dimension of the transform result of the sample is same as the number of the used transforming axes.

Conventional LPP may suffer from the following problem: in the case of the so-called SSS problem where the dimension of the sample is larger than the number of the samples, the eigenequation (1) cannot be directly solved since the matrix D is usually singular. Indeed, it is clear that the rank of D_1 must not be larger than the number of the samples. Consequently, if the dimension of the sample vector is larger than the number of the samples, (1) cannot be solved directly. On the other hand, if one modifies Eq. (1) into the form following form:

$$(D_1 + \mu I)^{-1} L_1 z = \lambda z \tag{2}$$

where μ is a small positive constant, conventional LPP solution can be directly worked out by solving (2). In practice, a similar procedure to solve singular eigenequation has been widely used in numerical computation [17], [18]. However, another problem arises in the above procedure for solving LPP. It is when the obtained minimum eigenvalue solution serves as LPP transforming axes of the SSS problem, the transform result will have poor data representation. This is because in the case of the SSS problem, there are a number of zero eigenvalues and conventional LPP will take as transforming axes the eigenvectors corresponding to zero eigenvalues of the eigenequation. Consequently, as shown in the following theorem, different samples may produce the same transform result.

Theorem 1. In the case where the minimum eigenvalue of the eigenequation is zero, after conventional LPP transforms samples into new representations by using the eigenvectors corresponding to the minimum eigenvalues as transforming axes, a sample will statistically have the same representation as its 'neighbors'.

Proof. Based on (2), we have $z^T L_1 z = \lambda z^T (D_1 + \mu I) z$. If (2) has zero eigenvalues and z_0 is the eigenvector corresponding to a zero eigenvalue, then we can obtain $z_0^T L_1 z_0 = 0$. We know that $\frac{1}{2}\sum_{ij}(y_i - y_j)^2 w_{ij} = z^T L_1 z$ [4]. As a result, if the eigenvector corresponding to a zero eigenvalue is taken as the transforming axis, the transform result will statistically satisfy the following condition $\frac{1}{2}\sum_{ij}(y_i - y_j)^2 w_{ij} = 0$, which implies that in the transform space the 'neighboring' samples must have the same representation. Indeed, because $(D_1 + \mu I)^{-1} L_1$ is positive semi-definitive, no negative eigenvalue exists and the minimum eigenvalue is of course zero. Theorem 1 shows that the conventional LPP solution is not quite subject to the purpose of local structure preservation, which requires that the transform results of 'neighbor samples' be also close rather than the same. Another drawback of conventional LPP is that the 'neighbor samples' from different same classes also statistically have the same representation in the transform space. We call this problem the presentation confusion problem, because it is disadvantageous for pattern classification tasks.

3 A New LPP Solution Scheme

3.1 Formal Description of the New Solution Scheme

In this subsection, we present a new LPP solution scheme that does not suffer from the same difficulties as conventional LPP. Suppose that $\alpha_1, \alpha_2, ..., \alpha_r$ are the eigenvectors corresponding to positive eigenvalues of D_1 and $\alpha_{r+1}, \alpha_{r+2}, ..., \alpha_N$ are the eigenvectors corresponding to the zero eigenvalues. It is easy to prove that valid LPP solution should be from the range space of D_1. Therefore, the new solution scheme is designed as follows. We first define that $R = [\alpha_1 \quad \alpha_2 ... \alpha_r]$. Using R, we respectively transform D_1 and L_1 into the following matrices: $\bar{D} = R^T D_1 R$, $\bar{L} = R^T L_1 R$. We construct the eigenequation $\bar{L}\bar{z} = \lambda \bar{D}\bar{z}$ (3) Since $\bar{D}$ is of full rank, we can directly solve this equation. Moreover, according to experimental results, the rank of $\bar{L}$ is usually close to that of $\bar{D}$. As a result, (3) almost has no zero eigenvalue. Therefore, the minimum eigenvalue solution of (3) usually transforms close samples into close rather than the same presentation in the transform space. This means that the new solution scheme is more subject to the goal of LPP than conventional LPP. Let $\beta_1, \beta_2, ..., \beta_r$ denote the eigenvectors corresponding to increasing eigenvalues $\lambda_1, \lambda_2, ..., \lambda_r$ of Eq.(3). If the sample is required to be transformed into an m-dimensional space, we should select $\beta_1, \beta_2, ..., \beta_m$ as the m transforming axes of LPP.

3.2 Transform Procedure

We summarize the transform procedure associated with the new LPP solution scheme as follows. Using the matrix R, we first produce $x_1 = x^T R$ for a sample x. Then we transform x_1 into y, where $y = x_1 \mathrm{H}$, $\mathrm{H} = [\beta_1 \quad \beta_2 ... \beta_m]$.

4 Experimental Design and Results

In this section, we test the new solution scheme (NSS) and the conventional LPP solution. The experiment is performed on the ORL database (http://www.cam-orl.co.uk). In this experiment the first five face images of all the subjects are used as training samples, and the corresponding remaining images are regarded as test samples. Conventional LPP is implemented using three schemes. The first scheme (FS) takes the minimum eigenvalue solutions of (2) as the LPP solution. The second scheme (SS) takes nonzero minimum eigenvalue solutions of (2) as the LPP solution. The third scheme (TS) first resizes each image and then solves Eq. (1) as previous literatures did. To obtain non-singular D_1 and make Eq. (1) be directly solvable, we reduce the size of each image to 7 by 23 by down-sampling [19]. Fig. 1 shows that NSS obtains a low classification error rate with the minimum being 17.5%, whereas the three schemes of conventional LPP have higher error rates. This clearly demonstrates that our solution to unsupervised LPP outperforms the conventional LPP solution. The reason why FS performs badly is that in the SSS problem there are a number of zero eigenvalue serving as the minimum eigenvalues and consequently

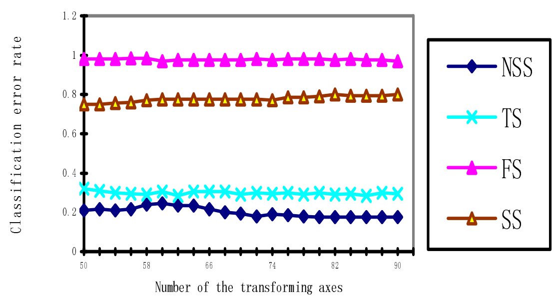

Fig. 1. Classification error rates of NSS, FS, SS and TS. The parameter k in the weight matrix W is set to $k = 1$.

the transform results of neighboring samples are the same even if they belong to different classes. This prevents the classification procedure from obtaining a high accuracy. Though SS and TS almost do not suffer from the zero-eigenvalue problem and outperform FS, their performances are still unsatisfactory. The following are the reasons. In SS, the minimum eigenvalues are quite small so that the transform results of two neighboring samples will be too close even if they are from different classes. Consequently, in SS weak 'presentation confusion' problem also occurs. For TS, the reason why it also does not obtain good result is that resizing images causes too much image information loss.

5 Conclusion

Solid theoretical analysis presented in this paper shows that the proposed LPP solution scheme is quite suitable for the SSS problem, because it does not suffer from the difficulties of conventional LPP such as the matrix singularity problem and the presentation confusion problem. The proposed scheme produces a well defined eigenequation that is directly solvable. Moreover, the obtained solution is more subject to the essence of locality preserving projection. Experimental results sufficiently show the validness and effectiveness of the solution proposed in this paper.

Acknowledgment

This work was partially supported by NSFC under Grant Nos.60602038, 60620160097, the National High-Tech Research and Development Plan of China (863) under Contract No.2007AA01Z195 as well as NSF of Guangdong Province, China under grant No.06300862.

References

1. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Proceedings of the Conference on Advances in Neural Information Processing Systems (2001)
2. Yang, J., Zhang, D., Yang, J.-Y., Niu, B.: Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics. IEEE Trans. Pattern Anal. Mach. Intell. 29(4), 650–664 (2007)
3. He, X., Niyogi, P.: Locality preserving projections. In: Proceedings of the Conference on Advances in Nerual Information Processing Systems (2003)
4. He, X., Yan, S., et al.: Face recognition using laplacianfaces. IEEE Trans. PAMI 27(3), 328–340 (2005)
5. Min, W., Lu, K., He, X.: Locality preserving projection. Pattern Recognition J. 37(4), 781–788 (2004)
6. He, X., Cai, D., Min, W.: Statistical and computational analysis of locality preserving projection. In: Proceedings of the 22nd International Conference on Machine Learning, pp. 281–288 (2005)

7. He, X., Yan, S., Hu, Y., Zhang, H.: Learning a Locality Preserving Subspace for Visual Recognition. In: ICCV 2003, pp. 385–393 (2003)
8. Zheng, Z., Yang, F., Tan, W., Jia, J., Yang, J.: Gabor feature-based face recognition using supervised locality preserving projection. Signal Processing 87, 2473–2483 (2007)
9. He, X.: Incremental semi-supervised subspace learning for image retrieval. ACM Multimedia, 2–8 (2004)
10. Cheng, J., Liu, Q., Lu, H., Chen, Y.-W.: Supervised kernel locality preserving projections for face recognition. Neurocomputing 67, 443–449 (2005)
11. Chen, H.-T., Chang, H.-W., Liu, T.-L.: Local discriminant embedding and its variants. In: Proc. CVPR (2005)
12. Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., Lin, S.: Graph embedding and extension: A general framework for dimensionality reduction. IEEE Transactions on PAMI 29(1) (2007)
13. Feng, G., Hu, D., Zhang, D., Zhou, Z.: An alternative formulation of kernel LPP with application to image recognition. Neurocomputing 69(13-15), 1733–1738 (2006)
14. Hu, D., Feng, G., Zhou, Z.: Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition. Pattern Recognition 40(1), 339–342 (2007)
15. Cai, D., He, X.: Orthogonal locality preserving indexing. In: SIGIR 2005, pp. 3–10 (2005)
16. Cai, D., He, X., Hu, Y., Han, J., Huang, T.S.: Learning a Spatially Smooth Subspace for Face Recognition. In: CVPR 2007 (2007)
17. Xu, Y., Zhang, D., Jin, Z., Li, M., Yang, J.-Y.: A fast kernel-based nonlinear discriminant analysis for multi-class problems. Pattern Recognition 39(6), 1026–1033 (2006)
18. Xu, Y., Yang, J.-y., Yang, J.: A reformative kernel Fisher discriminant analysis. Pattern Recognition 37(6), 1299–1302 (2004)
19. Xu, Y., Jin, Z.: Down-sampling face images and low-resolution face recognition. In: The third international conference on innovative computing, information and control (in Press)