Discriminant Locality Preserving Projections Based on Neighborhood Maximum Margin

Kezheng Lin, Youhu Rong, Di Wu, Liangwei Zhuang and Peng Li

Department of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China
link@hrbust.edu.cn, daboer923@163.com

Abstract

This paper took a research about the small size sample problem of the discriminant locality preserving projections method, and proposed the discriminant locality preserving projections method based on neighborhood maximum margin (NMMDLPP). Firstly, the training sample structured a weighted of K-nearest neighbor graph, and gave the weight parameter to each side of the nearest neighbor graph for obtaining the intraclass neighbors and interclass neighbors local geometry information of each point; then reduce the interval between the intraclass neighbors and increase the interval between the interclass neighbors with the result of transfer matrix, and applied the neighbor point optimal refactoring coefficient of the data to the objective function. This method chose the difference between the locality preserving between-class scatter and the locality preserving within-class scatter as the objective function to avoid of calculating the inversion of matrix. This method has conducted an experiment on the UMIST face database and Yale face database. Experimental results show that the NMMDLPP algorithm is superior to other algorithms in recognition rate. The recognition rate can reach more than 91.4%.

Keywords: feature extraction; face recognition; K-nearest neighbor rule; discriminant locality preserving projection

1. Introduction

In recent years, dimensionality reduction has been widely concerned in many fields such as machine learning, data mining and pattern recognition [1]. The aim is to map the high dimensional data into a low dimensional subspace, and keep the manifold structure of high dimensional data essential to subspace. Among the dimensionality reduction techniques, the most well-known techniques are the principal component analysis (PCA), linear discriminant analysis (LDA), locality preserving projections (LPP), and neighborhood preserving embedded (NPE) [2-3]. Although these methods have different suppositions, they can be put into a unified graph embedding framework with different constraints [4]. PCA is an unsupervised method, which does not take the class information into account [5]. The aim of LDA is to optimize the discriminate patterns of different classes by searching the projection axes, which the data points of different classed are far from each other, while constraining the data points of the same to be as close to each other as possible [6-7]. Among them, LPP is one of the most representative algorithm, there were many improved algorithm based on LPP theory [8]. Inspired by the LPP algorithm, someone has put forward the discriminant locality preserving projection (DLPP) algorithm, but there are still remained small size sample problem and information redundancy problem [9]. The categories of information data has not been fully used, which leads to a decrease in classification performance in pattern recognition problem of high-dimensional data [10]. Therefore, in view of the above questions, we propose

the discriminant locality preserving projections method based on neighborhood maximum margin (NMMDLPP) algorithm. It utilizes the sample data of local geometric information and class information to model the intraclass neighborhood scatters and interclass neighborhood scatters.

2. Discriminant Locality Preserving Projections

Discriminant locality preserving projections is introduced into the discriminant analysis, combined with Fisher criterion and locality preserving projections [11-12]. A sample set $X = \{x_1, x_2, \cdots, x_n\} \in R^{m \times n}$ for training is set for experiments. Each face image x_i belongs to one of the C face classes $\{X_1, \cdots, X_C\}$. Suppose the parameter n_c is the number of samples in the cth class, m_i and m_j is the mean projected vector for the ith class and j^{th} class respectively. The parameter y_i^c represents the ith projected vector in the cth class. Therefore, the objective function of DLPP is as follows:

$$\frac{\sum_{i,j=1}^{C} (m_i - m_j) B_{ij} (m_i - m_j)^T}{\sum_{c=1}^{C} \sum_{i,j=1}^{n_c} (y_i^c - y_j^c) W_{ij}^c (y_i^c - y_j^c)^T} \tag{1}$$

Where the parameter m_i is the equation of $m_i = \frac{1}{n_i} \sum_{k=1}^{n_i} y_k^i$, m_j is the equation of $m_j = \frac{1}{n_j} \sum_{k=1}^{n_j} y_k^j$, the parameter n_i and n_j is the number of samples in the ith and j^{th} class respectively. The parameter W_{ij}^c is the weight between y_i^c and y_j^c, $W_{ij}^c = \exp(-\|x_i^c - x_j^c\|^2 / \sigma^2)$, $B_{ij} = \exp(-\|f_i - f_j\|^2 / \sigma^2)$ is the weight between m_i and m_j. Suppose $D_i (i = 1, 2, \cdots, k)$ is a diagonal matrix, $A = [a_1, a_2, \cdots, a_k]$ is a transformation matrix, L and H are Laplacian matrices, and D_i entries are column (or row) sum of W_c. $W = \begin{bmatrix} W_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & W_c \end{bmatrix}$; $D = \begin{bmatrix} D_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & D_k \end{bmatrix}$; $D_{ii}^c = \sum_j W_{ij}^c$; $W_{ij}^c = \exp(-\|x_i^c - x_j^c\|^2 / t)$.

The f_i is the mean of the ith class, that is, $f_i = (1/n_i) \sum_{k=1}^{ni} x_k^i$. So, that is, $Y = A^T X$. The objective function can be simplified into:

$$J_1(A) = \frac{A^T F H F^T A}{A^T X L X^T A} \tag{2}$$

Where L=D-W, H=E-B, E is a diagonal matrix, and its elements are column (or row) sum of B, $E_{ii} = \sum_j B_{ij}$; $F = [f_1, f_2, \cdots, f_c]$. Suppose that the locality preserving within-class scatter is $S_w^L = XLX^T$, the locality preserving between-class scatter is $S_b^L = FHF^T$. The matrices XLX^T and FHF^T are symmetric and positive semi-definite. So, as for the definition, we can see the locality preserving total scatter $S_t^L = S_b^L + S_w^L$. The

transformation matrix can be obtained by the objective function (2). To solve the generalized eigenvalue problem.

$$(FHF^T)a_i = \lambda_i(XLX^T)a_i, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \tag{3}$$

or

$$FHF^T a_i = \lambda_i XLX^T a_i, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \tag{4}$$

Therefore, based on the above analysis, the optimal projection each column of the matrix as feature vector are $(S_w^L)^{-1}S_b^L = (XLX^T)^{-1}MHM^T$; and $rank(XLX^T) \leq rank(L) \leq n < m$. Thus, the locality preserving between-class scatter matrix S_w^L is singular. Therefore, it cannot obtain the inverse matrix $(S_w^L)^{-1}$ of locality preserving between-class scatter.

3. Discriminant Locality Preserving Projections based on Neighborhood Maximum Margin

The method first begins to build an adjacent graph ζ by k-neighborhood for all points. Suppose training samples $X = \{x_1, x_2, \cdots, x_n\}$, $x_i \in R^d$ are a group of high dimensional data, the purpose of NMMDLPP is to find the optimal projection matrix A, $A \in R^{d \times D}$, $D \,\square\, d$, $A = [a_1, a_2, \cdots, a_D]$, making the original high-dimensional data projected into the lower space to get a set of corresponding low-dimensional feature data $Y = \{y_1, y_2, \cdots, y_n\}$, $y_i \in R^D$, $y_i = A^T x_i$. For a data point x_i, let ω_{ij} be the set of k-nearest neighbors of it. The objective function of neighborhood maximum margin criterion is defined as follows:

$$J = \max\left(\sum_{i=1}^{n}\sum_{y_i \in N_i^w}\left\|\omega_{ij}^b(y_i - y_j)\right\|^2 - \sum_{i=1}^{n}\sum_{y_k \in N_i^w}\left\|\omega_{ik}^w(y_i - y_k)\right\|^2\right) \tag{5}$$

Where the parameter w_{ij}^b is the refactoring coefficient of between-class and the parameter w_{ij}^w is the refactoring coefficient of within-class. Let N_i^b denote the intraclass neighbors in the k-neighborhood ω_{ij} (*i.e.,* neighbors from the same class as x_i) and N_i^w the interclass neighbors of x_i in ω_{ij} (*i.e.,* neighbors from different classes).

$$\omega_{ij}^b = \begin{cases} \dfrac{\exp(\dfrac{-\|x_i - x_j\|^2}{t})}{\sum_{j=1}^{K_b}\omega_{ij}^b} & x_j \in N_i^b \\ 0 & otherwise \end{cases}, \omega_{ik}^w = \begin{cases} \dfrac{\exp(\dfrac{-\|x_i - x_j\|^2}{t})}{\sum_{k=1}^{K_w}\omega_{ik}^b} & x_k \in N_i^w \\ 0 & otherwise \end{cases}$$

Where the parameter t is a constant, K_w is number of within-class adjacent points and K_b is number of between-class adjacent points. According to formula $y_i = A^T x_i$, we can get the following equation:

$$\sum_{i=1}^{n}\sum_{y_j \in N_i^b}\left\|\omega_{ij}^b(y_i - y_j)\right\|^2 = tr[\sum_{i=1}^{n}\sum_{y_j \in N_i^b}\omega_{ij}^b(y_i - y_j) \times \omega_{ij}^b(y_i - y_j)^T]$$

$$= tr[A^T\sum_{i=1}^{n}\sum_{x_j \in N_i^b}\omega_{ij}^b \times \omega_{ij}^b(x_i - x_j)(x_i - x_j)^T A] = A^T tr(S_b^L)A \quad (8)$$

$$\sum_{i=1}^{n}\sum_{y_k \in N_i^w}\left\|\omega_{ik}^w(y_i - y_k)\right\|^2 = tr[\sum_{i=1}^{n}\sum_{y_k \in N_i^w}\omega_{ik}^w(y_i - y_k) \times \omega_{ik}^w(y_i - y_k)^T]$$

$$= tr[A^T\sum_{i=1}^{n}\sum_{x_k \in N_i^w}\omega_{ik}^w \times \omega_{ik}^w(x_i - x_k)(x_i - x_k)^T A] = A^T tr(S_w^L)A \quad (9)$$

Where the parameter S_b^L is matrix of locality preserving between-class scatter and S_w^L is matrix of locality preserving within-class scatter. They are defined as follows:

$$S_b^L = \sum_{i=1}^{n}\sum_{x_j \in N_i^b}\omega_{ij}^b \times \omega_{ij}^b(x_i - x_j)(x_i - x_j)^T, \quad S_w^L = \sum_{i=1}^{n}\sum_{x_k \in N_i^w}\omega_{ik}^w \times \omega_{ik}^w(x_i - x_k)(x_i - x_k)^T$$

Suppose the parameter δ is a positive regulator. The affinity weight integrates the local weight, that is, $\exp(-\left\|x_i - x_j\right\|^2/\delta)$. The affinity weights for intraclass neighborhoods of all points and interclass neighborhoods of all points are defined as follows:

$$W_{ij}^+ = \begin{cases} \exp(-\dfrac{\left\|x_i - x_j\right\|^2}{\delta})(1 + \exp(-\dfrac{\left\|x_i - x_j\right\|^2}{\delta})) & x_j \in N_i^b \; or \; x_i \in N_i^b \\ 0 & otherwise \end{cases} \quad (12)$$

$$W_{ij}^- = \begin{cases} \exp(-\dfrac{\left\|x_i - x_j\right\|^2}{\delta})(1 - \exp(-\dfrac{\left\|x_i - x_j\right\|^2}{\delta})) & x_j \in N_i^w \; or \; x_i \in N_i^w \\ 0 & otherwise \end{cases} \quad (13)$$

According to (12) and (13), the formula $1 + \exp(-\left\|x_i - x_j\right\|^2/\delta)$ is intraclass discriminating weight, which can represent the class information of the same classes. The formula $1 - \exp(-\left\|x_i - x_j\right\|^2/\delta)$ is interclass discriminating weight. Therefore, after the projection locality preserving between-class scatter and locality preserving within-class scatter can be expressed as the following two formulas: $A^T S_b^L A$, $A^T S_w^L A$. Suppose the parameter k is a nonnegative constant which can balances the relative merits of maximizing the locality preserving between-class scatter and the minimization of the locality preserving within-class scatter. So the objective function of NMMDLPP is defined as follows:

$$J_L(A) = tr(A^T(S_b^L - kS_w^L)A) = tr(A^T(FHF^T - kXLX^T)A) \quad (14)$$

The objective function of NMMDLPP is based on the difference form and the matrix $S_b^L - kS_w^L$ is symmetric. Thus, the small size sample and the inverse matrix operation problem are avoided. Suppose the parameter F is a matrix $S_b^L - kS_w^L$. Figure 1 shows the flow chart of the NMMDLPP algorithm.
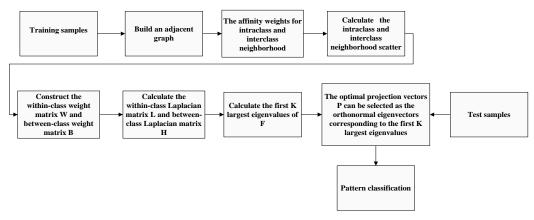
Figure 1. Flow Chart of the NMMDLPP Algorithm

So the criterion in (14) can be maximized by solving: $\max\limits_{v_i^T v_j = \delta_{ij}, i, j = 1, 2, \cdots g} \sum\limits_{t=1}^{g} v_t^T (S_b^L - k S_w^L) v$ .The matrix $S_b^L - k S_w^L$ eigenvalue decomposition can be obtained the first g largest eigenvalue $\lambda_1, \lambda_2, \cdots, \lambda_k$ and the orthonormal eigenvectors $v_1, v_2, \cdots, v_g$. So the vectors $v_1, v_2, \cdots, v_g$ are the optimal projection vectors, that is, $(S_b^L - k S_w^L) v_j = \lambda_j v_j$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_g$.

4. Experimental Results

This section will conduct an experiment about the effectiveness of the NMMDLPP method on UMIST and Yale face database. There are twenty face images in the UMIST face database, which are common and rich in gesture variation, as shown in Figure 2. The face image of the Yale face database has the variation of light and expression, the face expression view keeps the same with the all positive face image pixel of 320×243, as shown in Figure 3.
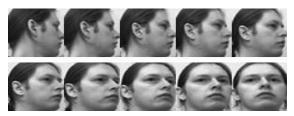

Figure 2. Images of One Person in UMIST

Figure 3. Images of One Person in Yale

4.1. Experiments on the UMIST Face Database

Take the face image from the UMIST face database as training sample set consisted of 6 faces image. The remaining sample image would be used as the test sample set of NMMDLPP method proposed in this paper. There are two experiments on the UMIST face database. The first experiment is as shown in Figure 4. Firstly, it researches the variation condition of the recognition rate of NMMDLPP with varying the k-neighborhood parameter w_k of the structure neighbor graph. The increase step is 2 with the ranges from 1 to 15. The second experiment is as shown in Figure 5. It surveys the comparison of the NMMDLPP method proposed in this paper and the method of PCA, LDA, LPP, DLPP with different feature dimension, when parameter w_k is 15. The general recognition performance of the recognition rate from the experiment result of the NMMDLPP method and the four methods on the UMIST face database is as shown in Table 1.

As the Figure 4 shows, the recognition rate with different parameter w_k increase rapidly with the feature dimension varied 0 to 5, and the recognition rate with different parameter w_k is very close with the feature dimension less than 5. Take the k-neighborhood parameter w_k of 7 as the boundary with the feature dimension more than 5, the recognition performance with w_k more than 7 is better than the recognition performance with w_k less than 7, the recognition rate with different parameter w_k increase slowly with the feature dimension varied 5 to 10. The recognition performance of method can be the best with the w_k varied from 13 to 15. As the Figure 5 and Table 1 show, when the feature dimension of NMMDLPP method is 10, the recognition rate is lower than the DLPP method within 3%, and is more than the recognition rate of the four methods with the other feature dimension. The recognition rate can reach the highest figure of 91.4% with the feature dimension of 14.
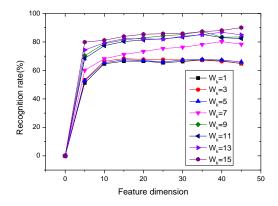

Figure 4. The Average Recognition Rates of NMMDLPP versus w_k on UMIST Face Database
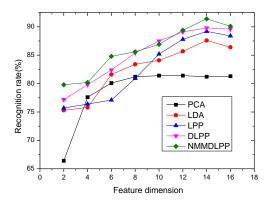
Figure 5. The Average Recognition Rates versus Feature Dimension on UMIST Face Database

Table 1. The Maximal Average Recognition Rates of Each Method on UMIST with the Feature Dimension

Method	Recognition rate (%)	Feature dimension
PCA	81.4	10,12
LDA	87.6	14
LPP	89.2	12
DLPP	89.8	16
NMMDLPP	91.4	14

4.2. Experiments on the Yale Face Database

The training sample set consist of 3 graphs of each face in the Yale face database experiment with the 45 samples of 3×15, and the remaining 120 images of 8×15 constitute the test sampling set. The NMMDLPP recognition rate varies with varying the k-neighborhood parameter w_k of the structured neighbor graph, see figure 6.When the parameter w_k is 15, the different method recognition rate with different feature dimension is as shown in Figure 7. The highest recognition rate resulted from Yale face database and the feature dimension is shown in Table 2.
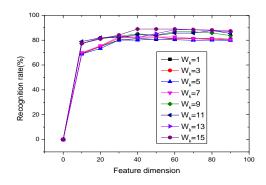

Figure 6. The Average Recognition Rates of NMMDLPP versus w_k on Yale Face Database
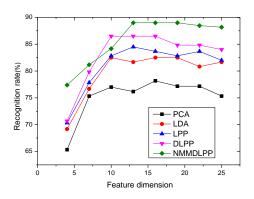
Figure 7. The Average Recognition Rates versus Feature Dimension on Yale Face Database

Table 2. The Maximal Average Recognition Rates of Each Method on Yale with the Feature Dimension

Method	Recognition rate(%)	Feature dimension
PCA	78.17	16
LDA	82.50	10
LPP	84.50	13
DLPP	86.50	10,13
NMMDLPP	89.00	16,19

As the Figure 6 shows, the recognition rate of NMMDLPP method increases rapidly with the different parameter w_k with the feature dimension varied from 0 to 10; the recognition rate increases slowly with the feature dimension varied from 10 to 30. When the parameter w_k is 15, the method can reach the highest recognition rate. As the Figure 7 and Table 2 shows, NMMDLPP method is greater than the other method recognition rate with the low feature dimension, and the recognition rate is lower than the recognition rate of DLPP method with the feature dimension varied from 6 to 13. When the feature dimension is more than 13, the recognition rate of NMMDLP method has a positive improvement than other recognition methods. The recognition rate of NMMDLP method can reach 89% with the feature dimension varied from 16 to 19.

5. Conclusion

This paper proposes the discriminant locality preserving projection based on neighborhood maximum margin. It improves the objective function, avoids the calculation of inversion and solves the small size sample problem effectively. The objective function introduced a parameter and balanced the maximum between-class and the minimum within-class for achieving the best optimization of recognition performance. Besides, the local weight of the neighbor image and discriminant weight could express the local neighbor structure and class information of the data, which can strength the class effect.

Acknowledgements

References

[1] H. Zhang, W. Deng, J. Guo and J. Yang, "Locality preserving and global discriminant projection with prior information", Machine Vision and Applications, vol. 21, no. 4, **(2010)**, pp. 577-585.

[2] J.-B. Li, Y. Yu, Z.-M. Yang and L.-L. Tang, "Breast tissue image classification based on semi-supervised locality discriminant projection with kernels", Journal of Medical Systems, vol. 36, no. 5, **(2012)**, pp. 2779-2786.

[3] F. Dornaika and A. Assoum, "Enhanced and parameterless locality preserving projections for face recognition", Neurocomputing, vol. 99, **(2013)**, pp. 448-457.

[4] X. Chen, W. Liu, J. Lai, Z. Li and C. Lu, "Face recognition via local preserving average neighborhood margin maximization and extreme learning machine", Soft Computing, vol. 16, no. 9, **(2012)**, pp. 1515-1523.

[5] C. Zhao, Z. Lai, D. Miao, Z. Wei and C. Liu, "Graph embedding discriminant analysis for face recognition", Neural Computing and Applications, vol. 24, nos. 7-8, **(2014)**, pp. 1697-1706.

[6] X. Li, Y. Pang and Y. Yuan, "L1-Norm-Based 2DPCA.. IEEE Transactions on Systems, Man, and Cybernetics", Part B, **(2010)**, pp. 1170 – 1175.

[7] Z. Fan, Y. Xu, D. Zhang, "Local linear discriminant analysis framework using sample neighbors", Neural Networks, IEEE Transactions on, vol. 22, no. 7, **(2011)**, pp. 1119-1132.

[8] Y. Hou, I. Song, H. K. Min and C. H. Park, "Complexity-reduced scheme for feature extraction with linear discriminant analysis", Neural Networks and Learning Systems, IEEE Transactions on, vol. 23, no. 6, **(2012)**, pp. 1003-1009 .

[9] X. He, S. Yan, Y. Hu, P. Niyogi and H.-J. Zhang, "Face recognition using Laplacianfaces", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, no. 3, **(2005)**, pp. 328-340.

[10] B. F. Klare, M. J. Burge, J. C. Klontz, R. W. Vorder Bruegge and A. K. Jain, "Face recognition performance: Role of demographic information. Information Forensics and Security", IEEE Transactions on, vol. 7, no. 6, **(2012)**, pp. 1789-1801.

[11] Y. Chen, Zhenzhen Li and Z. Jin, "Feature extraction based on maximum nearest subspace margin criterion", Neural Processing Letters, vol. 37, no. 3, **(2013)**, pp. 355-375.

[12] W. Yang, C. Sun, H. S. Du and J. Yang, "Feature extraction using Laplacian maximum margin criterion", Neural Processing Letters, vol. 33, no. 1, **(2011)**, pp. 99-110.