Int. J. Electron. Commun. (AEÜ) 69 (2015) 1724–1732

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

journal homepage: www.elsevier.com/locate/aeue

Local similarity preserving projections for face recognition

Pu Huang [a,*], Guangwei Gao [b]

[a] School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[b] Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

ARTICLE INFO

Article history:
Received 11 October 2013
Accepted 21 August 2015

Keywords:
Face recognition
Dimensionality reduction
Locality preserving projections (LPP)
Graph construction
Local similarity preserving projections (LSPP)

ABSTRACT

In graph embedding based learning algorithms, how to construct the local neighborhood graphs in applications is a difficult but important problem. In this paper, we propose a novel supervised subspace learning method called local similarity preserving projections (LSPP) for linear dimensionality reduction (DR). LSPP seeks to project the original high-dimensional data into a subspace, which preserves the local neighborhood structure of the data in a certain sense. Compared with most existing DR algorithms, such as locality preserving projections (LPP) which is unsupervised in nature and predefines the neighborhood parameters, LSPP takes special consideration of class information to guide the procedure of graph construction, which effectively avoids the difficulty of neighborhood parameter selection and shows more valuable discriminatory information for classification tasks. To evaluate the performance of LSPP, we conduct extensive experiments on three face databases, i.e. Yale, FERET and AR face datasets. The results corroborate that LSPP delivered promising performance compared with other competing methods such as PCA, LDA, LPP, Supervised LPP, LDP, SLPP and MFA.

1. Introduction

Recently, the appearance-based face recognition methods have aroused considerable interests in image processing and computer vision fields. Generally, these approaches treat each face image of size $p \times q$ as a point in $p \times q$ dimensional image space. In practice, however, these $p \times q$-dimensional spaces are too large to allow robust and fast recognition. Dimensionality reduction (DR) is an effective approach to deal with this problem. Over the past few decades, a variety of dimensionality reduction techniques including linear and nonlinear methods, supervised and unsupervised methods have been well developed. Among them, principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2] are two representative linear approaches.

PCA and LDA have been successfully applied to face recognition [3–5]. PCA seeks to project the original data into a low-dimensional subspace, which is spanned by the eigenvectors associated with the largest eigenvalues of the sample covariance matrix. PCA is guaranteed to produce a compact representation of the input data in the sense of minimizing mean squared error (MSE). However, PCA does not take the class information into account and thus may be not reliable for classification task. Different from PCA, which has nothing to do with the class information, LDA takes full consideration of class labels of the input data. LDA aims to maximize between-class scatter and simultaneously minimize within-class scatter. It is generally believed that the label information can improve the discriminative ability of recognition algorithms. Thus LDA can enhance class separability in comparison with PCA.

However, both PCA and LDA can see only the global Euclidean structure of the original data. If data points reside on a nonlinear sub-manifold, the two methods may fail to discover the intrinsic geometric structure of the data. In order to characterize those nonlinear data, several manifold learning methods with local linear but global nonlinear transformation are put forward, such as locally linear embedding (LLE) [6], Isometric mapping (ISOMAP) [7], and Laplacian Eigenmap [8]. These techniques do yield impressive visualization results on some benchmark data such as handwritten digits and facial images, whereas their implicit maps are defined only on the training data. Therefore, they might be unsuitable for feature extraction for pattern classification tasks. To overcome this limitation, He et al. extend Laplacian Eigenmap to its linearized version, i.e., locality preserving projections (LPP) [9–12] for an explicit map. LPP is also known as a linear graph embedding method by building a graph incorporating neighborhood information of the data set to preserve the local structure in the low dimensional space.

To improve the discriminantive ability of LPP, some supervised LPP methods [13–18] by combining locality and label information

* Corresponding author at: School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China. Tel.: +86 15295539600.
E-mail addresses: huangpu3355@163.com, huangpu@njupt.edu.cn (P. Huang).

http://dx.doi.org/10.1016/j.aeue.2015.08.009
1434-8411/

have been derived. However, the recognition capability for LPP and its supervised versions is limited due to that it is modeled based on a characterization of 'locality'. If the pattern needs to be classified resides in multi-manifolds and two or more modes have a common axis, then the locality preserving algorithms of manifold learning may result in overlapped embeddings belonging to different classes, which deteriorates the discrimination performance. To address such a problem, Yang et al. [19] proposed an unsupervised discriminant projection (UDP) method, which considers to minimize the local structure and simultaneously maximize the non-local structure. In addition, a number of methods [20–32], which combine the locality preserving technique and the linear discriminant analysis, have been developed to deal with this problem.

Furthermore, LPP suffers from the problem of neighborhood parameter selection. Since the neighborhood relationship is usually measured by an artificially constructed adjacent graph, graph construction has become an important issue. One of the most popular graph construction manners is based on the k nearest-neighbor. Once an adjacent graph is constructed, the edge weights are assigned by various strategies such as heat kernel or 0–1 way. Unfortunately, such an adjacent graph is artificially constructed in advance, thus it does not necessarily uncover the intrinsic local geometric structure of the samples. To make things worse, the performance of LPP is seriously sensitive to the neighborhood size k. More recently, several proposed LPP methods [33,34] have shown to be insensitive to k. On the other hand, some researches focus on how to construct the adjacent graph to avoid the selection of the neighborhood size k. Sparsity preserving projections (SPP) [35] aims to preserve the sparse reconstructive relationship of samples, which is achieved by constructing the adjacent graph by minimizing a L1 regularization-related objective function. Instead of predefining a same neighborhood size k for all samples, sample-dependent LPP (SLPP) [36] constructs the graph based on samples in question to determine the neighbors of each sample and similarities between sample pairs. However, like traditional LPP, when SLPP is applied to face recognition, it has several limitations such as the ignorance of class label information.

In this paper, a novel supervised dimensionality reduction method based on the idea of SLPP, namely local similarity preserving projections (LSPP) is proposed for face recognition. The novelty of LSPP over SLPP mainly comes from two aspects which are helpful for classification tasks: (1) the predefined similarities between two nodes can be adjusted according to their class information and show several good properties, and (2) the designed neighborhood decision rule can adaptively select the neighbors of each sample and guarantee that same-class samples are able to fall into the similarity neighborhood of a sample, which makes the algorithm more discriminative for classification.

The reminder of this paper is organized as follows. Section 2 gives a brief review of LPP and SLPP. Section 3 describes the proposed method LSPP. Section 4 explores theoretical connections of LSPP to some existing linear projection methods. Section 5 shows the experimental results and analysis, followed by the conclusions in Section 6.

2. Outline of LPP and SLPP

Given a set of n training data points $X = [x_1, \ldots, x_n]$, $x_i \in \mathrm{R}^N$, which belong to C classes $X_1, \ldots, X_C$. The generic form of linear dimensionality reduction is to project the high-dimensional data x_i into a low-dimensional space by a transformation matrix $A \in R^{N\times d}$ $(d < N)$ as: $y_i = A^{\mathrm{T}} x_i$. For the convenience of the following discussion, we denote $a \in R^N$ as the transformation vector. For the given data set X, a graph $G = \{V, W, E\}$ can be constructed, where V is the set of all data points in X, E is the set of edges connecting data points, and W is an adjacency matrix with weights characterizing the likelihood of point pairs.

2.1. LPP

LPP aims to find a projection that minimizes the local structure of the data in the transformed space. To fulfill such an objective, the criterion function of LPP can be formulated as:

$$J_{\mathrm{LPP}} = \frac{1}{2} \arg\min_a \sum_{ij} (y_i - y_j)^2 W_{ij} \tag{1}$$

where $y_i = a^{\mathrm{T}} x_i$ is the one-dimensional representation of x_i, and $W \in R^{N\times N}$ is a similarity matrix. A possible way of defining W is as follows:

$$W_{ij} = \begin{cases} \exp\left(\dfrac{-\left\|x_i - x_j\right\|^2}{2p^2}\right), & \text{if } \ x_i \in N_k(x_j) \ \text{ or } \ x_j \in N_k(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $\|\bullet\|$ is the Euclidean norm, $N_k(x_i)$ represents the set of k nearest neighbors of x_i, and p is the width parameter of the heat kernel.

By simple algebra formulation, the objective function can be reduced to

$$\begin{aligned} J_{\mathrm{LPP}} &= \frac{1}{2} \arg\min_a \sum_{ij} (a^{\mathrm{T}} x_i - a^{\mathrm{T}} x_j)^2 W_{ij} \\ &= \arg\min_a a^{\mathrm{T}} X (D - W) X^T a \\ &= \arg\min_a a^{\mathrm{T}} X L X^{\mathrm{T}} a \end{aligned} \tag{3}$$

where $L \in R^{N\times N}$ is the Laplacian matrix, $D \in R^{N\times N}$ is a diagonal matrix whose entries are column (or row, since W is symmetric) sum of W: $D_{ii} = \sum_j W_{ij}$.

To remove an arbitrary scaling in the embedding, a constant $a^{\mathrm{T}} X D X^{\mathrm{T}} a = 1$ is imposed. Then the minimization problem in Eq. (3) becomes

$$J_{\mathrm{LPP}} = \arg\min_{a^{\mathrm{T}} X D X^{\mathrm{T}} a = 1} a^{\mathrm{T}} X L X^{\mathrm{T}} a \tag{4}$$

The transformation vector a that minimize the projection is given by the minimum eigenvalue solution to the generalized eigenvalue problem:

$$X L X^{\mathrm{T}} a = \lambda X D X^{\mathrm{T}} a \tag{5}$$

2.2. SLPP

Different from LPP which uses the k nearest-neighbor strategy to determine neighborhood of samples, SLPP first utilizes an inequality with respect to the similarities between point pairs to extract neighbors of each sample, and if one sample is determined as a neighbor of another one, then the predefined similarity weights are assigned to connect the two data points. The weight matrix W^s of SLPP is defined by:

$$W_{ij}^s = \begin{cases} S_{ij}^s, & \text{if } \ S_{ij}^s > \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} S_{ik}^s \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $S_{ij}^s = \exp\left\{-\frac{d(x_i, x_j)}{2t^2}\right\}$ defines the similarity function, $d(x_i, x_j) = \left\|x_i - x_j\right\|^2 / \sum_{k=1}^{n} \left\|x_i - x_k\right\|^2$ is the distance between x_i and