Neural Networks 50 (2014) 33–46

Contents lists available at ScienceDirect

Neural Networks

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

Supervised orthogonal discriminant subspace projects learning for face recognition

Yu Chen *, Xiao-Hong Xu

College of Science, South China Agricultural University, Guangzhou, Guangdong, 510642, China

ARTICLE INFO

Article history:
Received 1 June 2013
Received in revised form 22 October 2013
Accepted 25 October 2013

Keywords:
Dimensionality reduction
Small sample size
Maximum margin criterion
Orthogonal constraint
Face recognition

ABSTRACT

In this paper, a new linear dimension reduction method called supervised orthogonal discriminant subspace projection (SODSP) is proposed, which addresses high-dimensionality of data and the small sample size problem. More specifically, given a set of data points in the ambient space, a novel weight matrix that describes the relationship between the data points is first built. And in order to model the manifold structure, the class information is incorporated into the weight matrix. Based on the novel weight matrix, the local scatter matrix as well as non-local scatter matrix is defined such that the neighborhood structure can be preserved. In order to enhance the recognition ability, we impose an orthogonal constraint into a graph-based maximum margin analysis, seeking to find a projection that maximizes the difference, rather than the ratio between the non-local scatter and the local scatter. In this way, SODSP naturally avoids the singularity problem. Further, we develop an efficient and stable algorithm for implementing SODSP, especially, on high-dimensional data set. Moreover, the theoretical analysis shows that LPP is a special instance of SODSP by imposing some constraints. Experiments on the ORL, Yale, Extended Yale face database B and FERET face database are performed to test and evaluate the proposed algorithm. The results demonstrate the effectiveness of SODSP.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Face recognition has attracted many researchers in the area of pattern recognition, machine learning, and computer vision because of its immense application potential over the past last decades (Belhumeur, Hespanha, & Kriengman, 1997; He, Zheng, & Hu, 2011; Jiang, Mandal, & Kot, 2008; Liu, Cheng, Yang, & Liu, 1992; Wang, Sung, & Yau, 2010; Wang, Xu, Zhang, & You, 2010; Zhao, Chellappa, Phillips, & Rosenfeld, 2003; Zheng, Lai, & Yuen, 2005). Numerous new techniques have been developed to handle different problems in face recognition, such as illumination and pose. The appearance based method is one of the most successful techniques. By using appearance based methods, an image is always represented by a high dimensional vector of pixels, which makes the storage space high and increases the computational cost. In addition, the high dimensionality also decreases the discrimination of face images. To overcome the curse of dimensionality, a natural way is to learn a subspace in which we can detect the reduced intrinsic dimension in the high dimensional image space. Principal component analysis (PCA) (Swets & Weng, 1996; Turk & Pentland, 1991) is a classical linear method for unsupervised subspace learning that transforms a data set consisting of a large number of interrelated variables to a new set of uncorrelated variables, while retaining most of the input data variations. Although PCA can generate the best approximations of the original samples, it does not take into account the separability between samples of different classes when extracting features. It is recognized that PCA may fail in capturing much useful discriminant information. Linear discriminant analysis (LDA) (Belhumeur et al., 1997) is another effective linear subspace algorithm for supervised learning. Differing from PCA, LDA computes a linear transformation which simultaneously maximizes the between-class scatter and minimizes the within-class scatter, achieving maximum discrimination. However, the intrinsic limitation of LDA is that its objective function requires the within-class covariance matrix to be nonsingular. But when the number of samples available is much smaller than the dimensionality of the sample space, LDA will suffer from the small sample size (SSS) problem.

Recently, many researchers pointed that large amounts of high-dimensional data probably lie on a nonlinear manifold (Cai, He, & Han, 2007; He, Cai, Yan, & Zhang, 2005; Kouropteva, Okun, & Pietikainen, 2003; Tenenbaum, de Silva, & Langford, 2000; Yan, Xu, Zhang, & Zhang, 2007). Linear models, such as PCA and LDA are particularly well suited for scenarios where the data originates from a subspace of the high-dimensional ambient space. On the other hand, data modeled by highly nonlinear manifolds cannot be well approximated by linear subspaces. To remedy this deficiency, a number of nonlinear dimension reduction techniques have been developed to learn the nonlinear structure of the manifold in

* Corresponding author. Tel.: +86 20 85280322.
E-mail addresses: chenyu204@scau.edu.cn, chenyu204@163.com (Y. Chen), xxh303@scau.edu.cn (X.-H. Xu).

0893-6080/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neunet.2013.10.006

the past few years, such as isometric feature (ISOMAP) (Belkin & Niyogi, 2003), locally linear embedding (LLE) (Roweis & Saul, 2000), and Laplacian Eigenmaps (LE) (Tenenbaum et al., 2000). This line of research is known as manifold learning, which explores the inherent nonlinear structure hidden in the observation space. However, one of the limitations in many existing manifold learning methods is the out-of-sample problem (Bengio, Paiement, & Vincent, 2003). And manifold learning, which is to pursue locality characterization of the data, is not originally and essentially designed for discrimination purpose. To overcome the out-of-sample problem, Neighborhood Preserving Embedding (NPE) (Wang et al., 2010) and Locality Preserving Projection (LPP) (He, Yang, & Hu, 2005) were proposed. LPP is essentially based on linear transformation that is actually performing linear dimension reduction, while considering manifold structure via adjacency graph. LPP and NPE aim to preserve local structure of data but differ in that NPE is a nonlinear learning for data embedding. Based on LPP, the Laplacianfaces algorithm (He, Yang et al., 2005) was further developed for face recognition.

The above manifold criterion is applied to Fisher criterion and marginal Fisher analysis (MFA) (Yan et al., 2007) as well as local Fisher discriminant analysis (LFDA) (Sugiyama, 2007), local discriminant embedding (LDE) (Chen, Chang, & Liu, 2005), local discriminating projection (LDP) (Zhao, Sun, Jing, & Yang, 2006), nonparametric discriminant analysis (NDA) (Li, Lin, & Tang, 2009), manifold discriminant analysis (MDA) (Wang & Chen, 2009), Locality Preserving Fisher Discriminant Analysis (LPFDA) (Zhao & Tian, 2009), Regularized Locality preserving Discriminant Embedding (RLPDE) (Pang, Andrew Beng, & Abas, 2012). However, it is also noted that those methods often suffer from a SSS problem.

As the orthogonal projection is of desirable property and often demonstrates good performance empirically, Ye (2006) proposed an orthogonal LDA (OLDA) algorithm, which is to solve the ratio trace optimization problem. Based on the LPP method, Cai, He, Han, and Zhang (2006) proposed a feature extraction approach named Orthogonal Laplacianface (OLPP) algorithm by constructing the orthogonal constraints. It builded an adjacency graph which can best reflect the geometry of the face manifold and the class relationship between the sample points. The optimal discriminant vectors in an iterative way were then obtained by preserving such a graph structure. However, OLDA and OLPP, based on Fisher criterion, also suffer from the singularity problem.

To address the high-dimensionality of data and the SSS problem, a novel linear subspace learning technique, called supervised orthogonal discriminant subspace projection (SODSP) is proposed in this paper. The points below highlight several aspects of our approach:

- SODSP considers the local and non-local information at the same time in designing the similarity matrix, which can explore the intrinsic structure of original data. Moreover, the class information is incorporated into the similarity matrix to model the manifold structure and enhance the recognition ability. The above two properties are united in SODSP by imposing an orthogonality constraint on the graph-based maximum margin criterion (MMC) (Li, Jiang, & Zhang, 2006), which draws the close samples closer while simultaneously separating the samples from different classes far enough.
- The proposed method requires the discriminant vectors to satisfy the orthogonality constraint, which ensures that the discriminant projection vectors are more powerful than the classical discriminant vectors in terms of discriminant ratio in a transformed space, and the orthogonality constraint can effectively preserve the metric structure of the data.
- Computationally, SODSP completely avoids the singularity problem incurred by many Fisher criterion based methods, since it seeks projections that maximizes the difference between local scatter and non-local scatter, which makes SODSP not involve any inverse matrix operation.
- In order to significantly alleviate the computational burden especially on high-dimensional data set, we develop an efficient and stable algorithm for performing SODSP.
- We theoretically analyze the relationship between LPP and SODSP, which shows LPP can be derived from SODSP by imposing some constraints.

The rest of this paper is organized as follows. In Section 2, we provide a brief introduction of LDA and MMC. The supervised orthogonal discriminant subspace projection (SODSP) algorithm is proposed in Section 3. Extensive experimental results on face recognition are presented in Section 4 and the conclusion is presented in Section 5.

2. Related works

Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ be a data set of given D-dimensional vectors of face images. Each data point belongs to one of the c object classes $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_c\}$. And l_i is the class label of $\mathbf{x}_i$, $l_i \in \{1, 2, \ldots, c\}$, where c is the number of classes. The goal of the dimensionality reduction is to find a linear transformation $\mathbf{Q} \in \mathbb{R}^{D\times d}$ that maps each vector $\mathbf{x}_i$ $(i = i, 2, \ldots, N)$ in the D-dimensional space to a vector $\mathbf{y}_i$ in the lower d-dimensional space by $\mathbf{y}_i = \mathbf{Q}^T\mathbf{x}_i$.

2.1. Linear discriminant analysis (LDA)

LDA is a supervised dimensionality reduction technique which derives a projection basis that separates data points from different classes as far as possible and compresses data points from the same classes as compact as possible, thus achieving maximum class discrimination in the dimensionality-reduced space. In LDA, three scatter matrices, namely the within-class, between-class and total-scatter matrices are defined as follows (Georghiades, Belhumeur, & Kriegman, 2001; Kim & Kittler, 2005):

$$\mathbf{S}_b = \sum_{i=1}^{c} M_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T. \tag{1}$$

$$\mathbf{S}_w = \sum_{i=1}^{c}\sum_{\mathbf{x}\in\mathbf{X}_i}(\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T. \tag{2}$$

$$\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_w \tag{3}$$

where $\mathbf{m}_i$ denotes the mean of class i and $\mathbf{m}$ is the global data mean. The number of vectors in class $\mathbf{X}_i$ is denoted by M_i. LDA learns a matrix, $\mathbf{U}$, which maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix as follows

$$\mathbf{Q}_{\text{opt}} = \arg\max_{\mathbf{Q}} \frac{|\mathbf{Q}^T\mathbf{S}_b\mathbf{Q}|}{|\mathbf{Q}^T\mathbf{S}_w\mathbf{Q}|} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_d]. \tag{4}$$

The solution $\{\mathbf{q}_i | i = 1, 2, \ldots, d\}$ is a set of generalized eigenvectors of $\mathbf{S}_b$ and $\mathbf{S}_w$, i.e., $\mathbf{S}_b\mathbf{q}_i = \lambda_i\mathbf{S}_w\mathbf{q}_i$. Usually, PCA is performed first to avoid the singularity problem which the within-class scatter matrix commonly encountered in face recognition (Swets & Weng, 1996; Turk & Pentland, 1991; Zhao, Chellappa, & Nandhakumar, 1998).

2.2. Maximum margin criterion (MMC)

MMC aims to maximize the average margins between different classes in the projected space. The objective function is given by

$$\mathbf{Q}_{\text{opt}} = \arg\max_{\mathbf{Q}} \mathbf{Q}^T(\mathbf{S}_b - \mathbf{S}_w)\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_d] \tag{5}$$

where $\mathbf{S}_w$ and $\mathbf{S}_b$ are, respectively, the between-class scatter matrix and the within-class scatter matrix as defined previously. The

Download English Version:

https://daneshyari.com/en/article/404012

Download Persian Version:

https://daneshyari.com/article/404012

Daneshyari.com