Neurocomputing 174 (2016) 756–766

Contents lists available at ScienceDirect

Neurocomputing

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

Facial expression recognition using sparse local Fisher discriminant analysis

Zhan Wang [a,b,*], Qiuqi Ruan [a,b], Gaoyun An [a,b]

[a] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
[b] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China

ARTICLE INFO

Article history:
Received 15 May 2015
Received in revised form
15 September 2015
Accepted 25 September 2015
Available online 9 October 2015

Keywords:
Facial expression recognition
Local Fisher discriminant analysis
Linearized Bregman iteration
Sparsity

ABSTRACT

In this paper, a novel sparse learning method, called sparse local Fisher discriminant analysis (SLFDA) is proposed for facial expression recognition. The SLFDA method is derived from the original local Fisher discriminant analysis (LFDA) and exploits its sparse property. Because the null space of the local mixture scatter matrix of LFDA has no discriminant information, we find the solutions of LFDA in the range space of the local mixture scatter matrix. The sparse solution is obtained by finding the minimum ℓ_1-norm solution from the LFDA solutions. This problem is then formulated as an ℓ_1-minimization problem and solved by linearized Bregman iteration, which guarantees convergence and is easily implemented. The proposed SLFDA can deal with multi-modal problems as well as LFDA; in addition, it has more discriminant power than LFDA because the non-zero elements in the basis images are selected from the most important factors or regions. Experiments on several benchmark databases are performed to test and evaluate the proposed algorithm. The results show the effectiveness of SLFDA.

1. Introduction

Facial expression recognition plays an important role in human-computer interaction and has attracted much attention in recent years. A face image is usually represented as a data point in high-dimensional space. Thus, dimensionality reduction becomes a fundamental process in recognition tasks. The goal of dimensionality reduction is to find a low-dimensional subspace that retains the useful information. The two most popular dimensionality reduction methods are principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2]. PCA is an unsupervised method that aims to find the optimal projection directions by maximizing the covariance over all the data samples. LDA, on the other hand, is a supervised method that finds the discriminant features by maximizing the ratio between the between-class and within-class scatters. However, in applications pertaining to pattern recognition, LDA usually suffers from the small sample size (SSS) problem [3], wherein the number of samples is much smaller than the dimensionality of the sample space. To address this problem, many variants of LDA such as Fisherface [4], null space LDA [5], LDA/QR [6], LDA/GSVD [7], have been proposed in the literature.

Both PCA and LDA can only reveal the global Euclidean structure; they fail to discover the underlying manifold structure. Recent studies show that the high-dimensional visual information in the real world lies on or is close to a smooth low-dimensional manifold. Therefore, many manifold learning based approaches have been proposed to discover the intrinsic manifold structure of the data, such as Isomap [8], Laplacian Eigenmap (LE) [9], locally linear embedding (LLE) [10] and local tangent space alignment (LTSA) [11]. However, these methods only yield mappings for training data. It remains unclear how to naturally evaluate maps obtained using the aforementioned methods for out-of-sample data. In order to tackle this problem, He and Niyogi proposed a new linear dimensionality reduction method called locality preserving projection (LPP) for face recognition [12]. Neighborhood preserving embedding (NPE) [13] is the linearization of LLE and preserves the local manifold structure as well as LPP. In essence, LPP and NPE attempt to linearly approximate the eigenfunctions of the Laplace-Beltrami operator L on the manifold. A linear version of the LTSA algorithm has also been proposed [14]. Shan et al. [15] compared several different linear subspace methods for facial expression recognition and concluded that supervised LPP outperforms other supervised methods.

On the other hand, in order to extract discriminant features that are important for image recognition and classification tasks, many manifold learning based discriminant analysis methods have been proposed. Yu et al. proposed discriminant LPP (DLPP) [16] to find a subspace that best discriminates different classes by

* Corresponding author.
E-mail addresses: wz6308@hotmail.com (Z. Wang), qqruan@bjtu.edu.cn (Q. Ruan), gyan@bjtu.edu.cn (G. An).

http://dx.doi.org/10.1016/j.neucom.2015.09.083
0925-2312/

maximizing the between-class distance, while minimizing the within-class distance. DLPP outperforms LPP for image feature extraction and representation. Lu et al. proposed DLPP based on the maximum margin criterion (DLPP/MMC) [17], which seeks to maximize the difference, rather than ratio, between the locality preserving between-class scatter and locality preserving within-class scatter. Yang et al. proposed a multi-manifold discriminant analysis (MMDA) method for image feature extraction [35]. Li et al proposed the discriminant LLE (DLLE) [18] for human gait recognition. In this method, the separability between different classes is enhanced by maximizing the margins between point pairs on different classes, while the local geometric structures within each class are presented by the LLE criterion. Zhang et al. proposed the discriminant orthogonal preserving projection (DONPP) [19], which takes into account both intraclass and interclass geometries. Sugiyama proposed a novel dimensionality reduction method called local Fisher discriminant analysis (LFDA) [20] for data visualization. It combines the ideas of Fisher discriminant analysis (FDA) [2] and LPP, and can work well when each class consists of several clusters. LFDA attempts to simultaneously maximize the between-class separability and preserve the within-class local structure. The projection matrix is spanned by the eigenvectors of a generalized eigenvalue problem.

The projection matrices of the above dimensionality reduction methods are not sparse. That is to say, each feature in the low-dimensional space is a linear combination of all the features of the original data and the coefficients of such a linear combination are generally non-zero, which makes the interpretation of the extracted features difficult especially when the data dimension is large. To overcome this drawback, Zou et al. proposed sparse PCA (SPCA) [21], which utilizes Lasso (or Elastic Net) to produce sparse principal components and improve performance for gene expression arrays analysis. Sparse discriminant analysis (SDA) [22] and sparse linear discriminant analysis (SLDA) [23] were proposed to learn a sparse discriminant subspace for feature extraction and classification in biological and medical data analysis. Both methods attempt to transform the original objective function into a regression-type problem, and add a Lasso penalty to obtain the sparse projection axis. One disadvantage of these two approaches is that the number of sparse vectors is at most $C-1$, where C is the number of classes. Lai et al. proposed sparse linear embedding [36], sparse 2D projections [37] and sparse tensor discriminant analysis (STDA) [38] for image feature extraction. These methods mainly utilize sparse regression, which is used in SPCA and SLDA, to find the sparse solution. if the regularization parameters are zeros, the above methods can derive the exact solutions of the original problem. However, these solutions are not sparse. If the parameters of the these l_1-norm penalty are larger than zero, these sparse solutions are the approximate solutions of the original objective function. Furthermore, These sparse learning methods enhance the generalization ability and provide the psychological and physiological interpretation.

In this paper, we proposed a novel facial expression recognition algorithm based on sparse LFDA (SLFDA). Because the null space of the local mixture scatter matrix defined in LFDA has no discriminant information, we remove it and find the solutions in the range space of the local mixture scatter matrix. The sparse solution is obtained by finding the minimum ℓ_1-norm solution from all the solutions of the LFDA. Finding the minimum ℓ_1-norm solution can be formulated as a ℓ_1-minimization problem, which can be solved by many numerical methods. The linearized Bregman method [24] is regarded as one of the most powerful methods for solving ℓ_1-minimization problem. Thus, we employ the linearized Bregman method to obtain the sparse projection vectors. The proposed SLFDA has three advantages: (1) Sparsity in SLFDA can control the weights of original variables or features. Therefore, the learn model can provide generalization ability better. (2) It can enhance the discriminant power of LFDA and work well for multi-modal problem. The projection vectors are sparse, which can make the physical meaning of the extracted features clear. Furthermore, the sparse projection matrix is a solution of SLFDA (3) As in LFDA; the dimension of the extracted features is not restricted as $C-1$. Moreover, the computational complexity is less than SLDA. The effectiveness of the proposed method is evaluated on two public facial expression databases and one face database that are commonly used by researchers who use pattern recognition methods for image recognition.

The rest of this paper is organized as follows: in Section 2, we introduce the problem of dimensionality reduction, and give an overview of LDA and LFDA. In Section 3, we present a detailed analysis of SLFDA. The experiments and corresponding discussion are given in Section 4. Finally, the conclusions are drawn in Section 5.

2. Related work

In this section, we present a brief review of previous work. The review consists of three parts: a description of the dimensionality reduction problem, an overview of LDA and an overview of LFDA algorithm.

2.1. The dimensionality reduction problem

The dimensionality reduction problem is described as follows: assume that we have a set of n samples: $x_1, x_2, \cdots, x_n, x_i \in \Re^m$, where $l_i \in \{1, 2, \cdots, C\}$ is the label of x_i,C is the number of classes, and $n = \sum_{k=1}^{C} n_k$,where n_k is the number of samples in the k th class. Let $X = [x_1, x_2, \cdots, x_n]$ be the data matrix, and assume that this data set has intrinsic dimension $d (d \ll m)$. That is to say, the points in data set X are lying on or near a manifold with dimension d that is embedded in m-dimensional space. The goal of dimensionality reduction is to find a mapping $F: \Re^m \to \Re^d, d \ll m : y_i = F(x_i), y_i \in \Re^d$. If the mapping is linear, it can be formulated as $y_i = V^T x_i$, where $V = [v_1, v_2, \cdots, v_d] \in \Re^{m \times d}$, and v_i is the projection vector. Hence the original n samples are mapped into a set of points $y_1, y_2, \cdots, y_d$ in d- dimensional space $\Re^d$.

2.2. LDA

LDA aims to find the discriminant features according to the Fisher criterion, that is, the within-class distance is minimized, simultaneously, the between-class distance is maximized. The within-class scatter matrix S_w and between-class scatter matrix S_b are respectively defined as follows:

$$S_w = \sum_{i=1}^{c} \sum_{x \in X_i} (x_i - m_i)(x_i - m_i)^T \tag{1}$$

$$S_b = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T \tag{2}$$

where m_i is the mean of the samples in class i and m is the mean of the total data. LDA seeks the optimal projection matrix by maximizing the following Fisher criterion:

$$V_{opt} = \arg\max_V \frac{V^T S_b V}{V^T S_w V} \tag{3}$$

The above optimization is equivalent to solving the following generalized eigenvalue problem:

$$S_b v_i = \lambda_i S_w v_i \tag{4}$$