Pattern Recognition 55 (2016) 207–214

Contents lists available at ScienceDirect

Pattern Recognition

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

L1-norm and maximum margin criterion based discriminant locality preserving projections via trace Lasso

Gui-Fu Lu [a,b,*], Jian Zou [a], Yong Wang [a]

[a] School of Computer and Information, AnHui Polytechnic University, WuHu, AnHui 241000, China
[b] Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information, Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China

ARTICLE INFO

Article history:
Received 9 April 2015
Received in revised form
18 January 2016
Accepted 28 January 2016
Available online 4 February 2016

Keywords:
Discriminant locality preserving projections
Feature extraction
Maximum margin criterion
L1-norm
Trace Lasso
L2-norm

ABSTRACT

Discriminant locality preserving projections based on maximum margin criterion (DLPP/MMC) is a useful feature extraction method since it has shown good performances in pattern recognition. The conventional DLPP/MMC, however, is not robust to noises and outliers since its objective function is based on L2-norm. In this paper, we propose a novel L1-norm and maximum margin criterion based discriminant locality preserving projections via trace Lasso (DLPP/MMC-L1TL). L1-norm rather than L2-norm is used in the formulation of DLPP/MMC-L1TL, which makes it be robust to noises and outliers. Besides, in order to improve the performance of DLPP/MMC-L1TL further, we use trace Lasso to regularize the basis vectors. Trace Lasso, which can balance L1-norm and L2-norm and consider sparsity and correlation of data simultaneously, is a recently proposed norm. An iterative procedure for solving DLPP/MMC-L1TL is also proposed in this paper. The experiment results on some data sets demonstrate the effectiveness of DLPP/MMC-L1TL.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Feature extraction plays a core role in many machine learning, pattern recognition problems [1–6] and a lot of feature extraction methods have been presented in the literature [7–10]. Among the various approaches, principal component analysis (PCA) [1,11] and linear discriminant analysis (LDA) [1,11,12] may be two of the most well-known ones.

PCA is an unsupervised feature extraction method since it does not consider any class label information. LDA, however, is a supervised feature extraction method since it uses the class label information in constructing the within-class scatter and between-class scatter matrices. Unfortunately, LDA suffers from the so-called small sample size (SSS) problem when the number of the data points in the training set is lower than the dimensionality of the data points. To deal with the SSS problem, many LDA-based methods, e.g. PCA+LDA, which is in the context of face recognition also known as Fisherfaces [12], the null space-based LDA [13], complete LDA [14] and maximum margin criterion (MMC) [15–17], etc., have been proposed in the recent years.

PCA and LDA are both global linear techniques and do not consider the local geometrical structure among the data points. Recently, some manifold learning algorithms, e.g. locally linear embedding (LLE) [18], Isomap [19], Laplacian Eigenmaps (LE) [20], etc., which can discover the geometrical structure of the underlying manifold, have been proposed in the literature. Unfortunately, these nonlinear methods yield maps that are defined only on the training data and it is unclear how to evaluate the maps on the testing data [21]. To address this problem, some linearization versions of these methods have been proposed. For example, locality preserving projection (LPP) [21], neighborhood preserving projection (NPE) [22] and isometric projection (IsoProjection) [23] are, respectively, the linearization versions of LE, LLE and Isomap.

Recently, Yu et al. [24] proposed a discriminant locality preserving projections (DLPP) approach to improve the recognition performance of the LPP method by considering the class label information. Similar to LDA, DLPP also suffers from the SSS problem. To address the SSS problem, Lu et al. [25] proposed a discriminant locality preserving projections based on maximum margin criterion method, giving encouraging classification performance.

The distance criteria used in all the aforementioned methods, however, are based on L2-norm, which makes them be sensitive to noise and outliers since the square operation in L2-norm exaggerates their effects. Compared with L2-norm, L1-norm is generally more robust to outliers. Then, some researchers turn to utilize L1-

* Corresponding author. at: School of Computer and Information, AnHui Polytechnic University, WuHu, AnHui 241000, China.
E-mail address: luguifu_tougao@163.com (G.-F. Lu).

http://dx.doi.org/10.1016/j.patcog.2016.01.029
0031-3203/© 2016 Elsevier Ltd. All rights reserved.

norm rather than L2-norm to develop the robust feature extraction approaches [26–37].

L1-norm-based principal component analysis (PCA-L1) [26], which is computationally efficient and rotationally invariant in the embedding space, is a recently proposed robust feature extraction method. Experimental evidence on some face image databases demonstrates that PCA-L1 can obtain much lower reconstruction errors than other PCA-based methods. Motivated by PCA-L1, Li et al. [27] and Pang et al. [30], respectively, developed L1-norm-based 2DPCA (2DPCA-L1) and L1-norm-based tensor analysis (TPCA-L1). Similar to conventional PCA, L1-norm-based PCA is also unsupervised and does not utilize the class label information. Then, the projection vectors learned by L1-norm-based PCA are not optimal for classification problems. Recently, some L1-norm-based LDA (LDA-L1) methods [32,34–36] have proposed for robust discriminant analysis. These LDA-L1 methods replace L2-norm in the conventional LDA formulation with L1-one and maximize the ratio of the L1-norm-based between-class dispersion to the L1-norm-based within-class dispersion. Zhong et al. [37] proposed a robust version of DLPP based on L1-norm maximization (DLPP-L1) which maximize the ratio of the L1-norm-based locality preserving between-class dispersion to the L1-norm-based locality preserving within-class dispersion.

In this paper, we focus on the robust version of DLPP/MMC and propose a novel L1-norm and maximum margin criterion-based discriminant locality preserving projections via trace Lasso (DLPP/MMC-L1TL). First, we replace L2-norm in the formulation of conventional DLPP/MMC with L1-norm, which makes DLPP/MMC-L1TL be robust to noise and outliers. Then, we use trace Lasso [38–40] to regularize the basis vectors. Trace Lasso, which can balance L1-norm and L2-norm, and consider sparsity and correlation of data simultaneously, is a recently proposed norm. We also propose an efficient iterative algorithm for solving DLPP/MMC-L1TL. The experiments on publicly available data sets confirm the effectiveness of the proposed method.

The remainder of the paper is organized as follows. In Section 2, we review briefly the DLPP and DLPP/MMC methods. In Section 3, we propose the DLPP/MMC-L1TL approach, including its objective function and algorithmic procedure. We report the experiment results in Section 4. Finally, we conclude the paper in Section 5.

2. Outline of DLPP and DLPP/MMC

Let $X=\{\mathbf{x}_j^i, j=1,2,...,n_i; i=1,2,...,k\}\in R^{d\times n}$ be the given training samples, where $\mathbf{x}_j^i$ is the jth samples of the ith class, k is the number of the classes, n_i is the number of the samples of ith class, d is the dimensionality of the training samples and $n=\sum_{i=1}^{k} n_i$ is the number of the data set. The objective function of DLPP (termed as DLPP-L2) is defined as:

$$\arg\max_{\mathbf{w}} \frac{\mathbf{w}^T MHM^T \mathbf{w}}{\mathbf{w}^T XLX^T \mathbf{w}} \tag{1}$$

where $\mathbf{w}\in R^d$ is the optimal projection vector, $M=[\mathbf{m}_1,\mathbf{m}_2,...,\mathbf{m}_k]$, here $\mathbf{m}_i=(1/n_i)\sum_{j=1}^{n_i}\mathbf{x}_j^i$ is the mean of the ith class, L and H are both Laplacian matrices, $L=D-W, D=diag\{D_1,...,D_k\}$, D_i is a diagonal matrix and its elements are column (or row) sum of $W^{(i)}$, here $W^i=[W_{jc}^i](i=1,...,k; j,c=1,2,...,n_i), W_{jk}^i=\exp(-\|\mathbf{x}_j^i-\mathbf{x}_k^i\|^2/\sigma^2)$ and σ is an empirically determined parameter, $H=E-B$, E is a diagonal matrix and its elements are column (or row) sum of B, here $B=[B_{ij}](i,j=1,2,...,k)$ and $B_{ij}=\exp(-\|\mathbf{m}_i-\mathbf{m}_j\|^2/\sigma^2)$.

The optimal projection vector $\mathbf{w}$ of DLPP can be obtained by solving the generalized eigenvalue problem:

$$(MHM^T)\mathbf{w}=\lambda(XLX^T)\mathbf{w} \tag{2}$$

When the SSS problem occurs, DLPP can not obtain its optimal projection vectors using Eq. (2) since MHM^T is singular. To address this problem, maximum margin criterion-based DLPP (DLPP/MMC) (termed as DLPP/MMC-L2) [25] has been developed. The formulation of DLPP/MMC is defined as follows:

$$\arg\max_{\mathbf{w}} \mathbf{w}^T MHM^T \mathbf{w} - \mathbf{w}^T XLX^T \mathbf{w} \tag{3}$$

The optimal projection vector $\mathbf{w}$ is the eigenvector which corresponds to the greatest eigenvalue of $MHM-XLX^T$.

3. L1-norm and maximum margin criterion-based discriminant locality preserving projections via trace Lasso

3.1. Problem reformulation

In this subsection, we will present our L1-norm and maximum margin criterion-based discriminant locality preserving projections via trace Lasso (DLPP/MMC-L1TL).

Let the eigenvalue decompositions of H and L, respectively, be

$$H=\begin{bmatrix} U_H & \overline{U}_H \end{bmatrix}\begin{bmatrix} D_H & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} U_H^T \\ \overline{U}_H^T \end{bmatrix} \tag{4}$$

and

$$L=\begin{bmatrix} U_L & \overline{U}_L \end{bmatrix}\begin{bmatrix} D_L & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} U_L^T \\ \overline{U}_L^T \end{bmatrix} \tag{5}$$

where $\begin{bmatrix} U_H & \overline{U}_H \end{bmatrix}\in R^{d\times d}$ and $\begin{bmatrix} U_L & \overline{U}_L \end{bmatrix}\in R^{d\times d}$ are both orthogonal matrices, $D_H\in R^{h\times h}$ and $D_L\in R^{l\times l}$ are both diagonal matrices and nonsingular, $h=\text{rank}(H)$ and $l=\text{rank}(L)$.

By using Eqs. (4) and (5), we can reformulate Eq. (3) as

$$\arg\max_{\mathbf{w}} \mathbf{w}^T MU_H D_H U_H^T D^T \mathbf{w} - \mathbf{w}^T XU_L D_L U_L^T X^T \mathbf{w} \tag{6}$$

Eq. (6) can also be rewritten as

$$\arg\max_{\mathbf{w}} \|\mathbf{w}^T MU_H D_H^{0.5}\|_2^2 - \|\mathbf{w}^T XU_L D_L^{0.5}\|_2^2 \tag{7}$$

where $\|\cdot\|_2$ denotes L2-norm. From Eq. (6) we can know that the objective function of DLPP/MMC is based on L2-norm. As mentioned in Section 1, L2-norm is much more sensitive to outliers than L1-norm. Thus we adopt L1-norm instead of L2-norm to design the objective function of the proposed method. Besides, in order to improve performance of the proposed method further, we use trace Lasso to regularize the basis vector $\mathbf{w}$. Motivated by these ideas, we define the objective function of DLPP/MMC-L1TL as follows:

$$\arg\max_{\mathbf{w}} \|\mathbf{w}^T F\|_1 - \|\mathbf{w}^T G\|_1 - \lambda\|X^T \text{Diag}VM(\mathbf{w})\|_* \tag{8}$$

where $F=[\mathbf{f}_1,\mathbf{f}_2,...,\mathbf{f}_h]=MU_H D_H^{0.5}, G=[\mathbf{g}_1,\mathbf{g}_2,...,\mathbf{g}_l]=XU_L D_L^{0.5}$, $\|\bullet\|_1$ denotes L1-norm, $\|\bullet\|_*$ is the trace norm of a matrix, i.e., the sum of its singular values and Diag$VM(\bullet)$ denotes conversion of a vector into a diagonal matrix. Note that trace Lasso, i.e.,$\|X^T\text{Diag}VM(\mathbf{w})\|_*$, contains the data sample matrix X. From [39] we can know that trace Lasso can adaptively interpolate between L1-norm and L2-norm, which makes it consider sparsity and correlation of data simultaneously.

3.2. Optimization procedure for DLPP/MMC-L1TL

In this subsection, we will introduce how to find the optimal $\mathbf{w}$ of Eq. (8).

Download English Version:

https://daneshyari.com/en/article/530457

Download Persian Version:

https://daneshyari.com/article/530457

Daneshyari.com