Neural Networks 97 (2018) 127–136

Contents lists available at ScienceDirect

Neural Networks

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

Matrix exponential based discriminant locality preserving projections for feature extraction

Gui-Fu Lu *, Yong Wang, Jian Zou, Zhongqun Wang

School of Computer Science and Information, AnHui Polytechnic University, WuHu, AnHui 241000, China

ARTICLE INFO

Article history:
Received 18 October 2016
Received in revised form 21 July 2017
Accepted 28 September 2017
Available online 16 October 2017

Keywords:
Dimensionality reduction
Discriminant locality preserving projections
Matrix exponential
Linear discriminant analysis

ABSTRACT

Discriminant locality preserving projections (DLPP), which has shown good performances in pattern recognition, is a feature extraction algorithm based on manifold learning. However, DLPP suffers from the well-known small sample size (SSS) problem, where the number of samples is less than the dimension of samples. In this paper, we propose a novel matrix exponential based discriminant locality preserving projections (MEDLPP). The proposed MEDLPP method can address the SSS problem elegantly since the matrix exponential of a symmetric matrix is always positive definite. Nevertheless, the computational complexity of MEDLPP is high since it needs to solve a large matrix exponential eigenproblem. Then, in this paper, we also present an efficient algorithm for solving MEDLPP. Besides, the main idea for solving MEDLPP efficiently is also generalized to other matrix exponential based methods. The experimental results on some data sets demonstrate the proposed algorithm outperforms many state-of-the-art discriminant analysis methods.

1. Introduction

In recent years, high-dimensional data is often used in pattern recognition and computer vision fields (Gu & Sheng, 2016; Gu, Sheng, Tay, Romano, & Li, 2015a; Gu, Sheng, Wang, Ho, Osman, & Li, 2015b; Gu, Sun, & Sheng, 2016; Wen, Shao, Xue, & Fang, 2015; Zhou, Wang, Wu, Yang, & Sun, 2016). Dimensionality reduction, which aims at transforming the high-dimensional data into the meaningfully low-dimensional ones, has been a key issue in these domains. In the past decades, a lot of dimensionality reduction methods have been developed and the most famous methods may be principal component analysis (PCA) (Fukunaga, 1990) and linear discriminant analysis (LDA) (Duda, Hart, & Stork, 2000).

PCA is an unsupervised method which aims to find a set of orthonormal basis vectors on which the variance over all the data is maximized. PCA is very popular since it is optimal in terms of minimum reconstruction error. However, PCA may be not suitable for classification problem since it does not use any class information in computing the basis vectors.

Different from PCA, LDA is a supervised technique for dimensionality reduction. LDA tries to find the most discriminant projection vectors by maximizing the between-class scatter matrix and minimizing the within-class scatter matrix simultaneously. In many pattern recognition tasks, however, LDA suffers from the well-known small sample size (SSS) problem where the number of samples is less than the dimension of samples.

In order to address the SSS problem, a lot of LDA-based methods have been proposed. PCA+LDA, also known as Fisherface (Belhumeour, Hespanha, & Kriegman, 1997), is a two-stage method, which first applies PCA to reduce the dimension of the high-dimensional data and then applies LDA for feature extraction. However, one potential problem for Fisherface is that some useful discriminant information may be lost in the PCA step. In Chen, Liao, Ko, and Yu (2000) proposed the null space LDA (NLDA) method, which projects the high-dimensional data onto the null space of the within-class scatter matrix where the between-class scatter matrix is maximized. Direct LDA (Yu & Yang, 2001) first removes the null space of the between-class scatter matrix and then computes the eigenvectors corresponding to the smallest eigenvalues of the reduced within-class scatter matrix. Maximum margin criterion (MMC), which is proposed by Li, Jiang, and Zhang (2004, 2006) and Song, Zhang, Mei, and Guo (2007), respectively, is another method to address the SSS problem encounter by LDA. In MMC, the inverse matrix of the within-class scatter matrix does not to be computed since the difference of between-class scatter matrix and within-class scatter matrix is used as discriminant criterion. Then the SSS problem is alleviated.

Though PCA and LDA are very useful for feature extraction, they can only preserve the global Euclidean structure and cannot discover the underlying manifold structure hidden in the high-dimensional data. To address the problem, some manifold learning methods have been developed, e.g., Isomap (Tenenbaum, Silva, &

* Corresponding author.
E-mail address: luguifu_tougao@163.com (G.-F. Lu).

https://doi.org/10.1016/j.neunet.2017.09.014
0893-6080/

Langford, 2000), locally linear embedding (LLE) (Roweis & Saul, 2000), and Laplacian Eigenmap (LE) (Belkin & Niyogi, 2003). Experiments demonstrate that these manifold learning methods are effective in finding the geometrical structure of the underlying manifold. However, it is difficult for these manifold learning methods to evaluate the map for unseen data. He, Yan, Hu, Niyogi, and Zhang (2005b) and He, Yan, Hu, and Zhang (2003) proposed the locality preserving projections (LPP) algorithm which is a linearization of LE and can provide a way to the projection of the new data samples. LPP can preserve the locality of data samples by building a graph incorporating neighborhood information among data samples. Similarly, neighborhood preserving embedding (NPE) (He, Cai, Yan, and Zhang, 2005a) is also a linearization of locally linear embedding (LLE). Although LPP is effective in discovering the manifold structure hidden in high-dimensional data, it is an unsupervised learning technique. Then, discriminant locality preserving projections (DLPP) (Yu, Teng, & Liu, 2006), which can utilize the class information, is proposed to address the problem of LPP. To overcome the limitation of LDA that it can only preserve the global geometry structure, Chen, Chang, and Liu (2005) proposed the local discriminant embedding (LDE) method. Yan, Xu, Zhang, Zhang, Yang, and Lin (2007) proposed a general framework, called graph embedding framework, to unify all the above-mentioned methods. Besides, based on the proposed framework, Yan et al. also proposed a new dimensionality reduction algorithm, called Marginal Fisher Analysis (MFA). Similar to LDA, the manifold learning methods also suffers from the SSS problem. Then, PCA is first used to reduce the dimension of samples and then uses these manifold learning algorithms for feature extraction. Similarly, some useful discriminant information may also be lost in the PCA step.

Zhang, Fang, Tang, Shang, and Xu (2010) proposed an exponential discriminant analysis (EDA) to address the SSS problem of LDA. In EDA, the matrix exponential is used, which makes the within-class and between-class scatter matrices be positive definite. The performance of EDA is good since no discriminant analysis is lost in performing the EDA algorithm. By borrowing the idea of EDA, Wang, Chen, Peng, and Zhou (2011) proposed the exponential LPP (ELPP) method, which aims at addressing the SSS problem of LPP. Similarly, Dornaika and Bosaghzadeh (2013) proposed the exponential local discriminant embedding (ELDE) algorithm to overcome the SSS problem of LDE without discarding any discriminant information. Furthermore, Dornaika and Traboulsi (2017) proposed the exponential semi-supervised discriminant analysis (ESDE), which is the semi-supervised extension of ELDE. Recently, Wang, Yan, Yang, Zhou, and Fu (2014) proposed an exponential graph embedding framework for dimensionality reduction. Unfortunately, all of these matrix exponential based methods are computationally intensive since the computational complexities are $O(d^3)$, where d denotes the dimension of samples and d is usually very large in many applications. More specifically, one has to compute the exponential of large matrices and a large matrix exponential eigenproblem, which makes these matrix exponential based methods to have high computational cost problem. So it is urgent to find new procedures for solving matrix exponential based methods efficiently.

In this paper, we propose a novel matrix exponential based discriminant locality preserving projections (MEDLPP), which can address the SSS problem of DLPP elegantly. Similar to other matrix exponential based methods, MEDLPP also suffers from the limitation that the computational complexity of MEDLPP is usually high. Then, we present an efficient algorithm for solving MEDLPP. The main idea for solving MEDLPP efficiently is also generalized to other matrix exponential based methods. The experimental results on some datasets demonstrate the proposed algorithm outperforms many state-of-the-art discriminant analysis methods.

The main contributions of this paper are as follows.

(1) We propose a matrix exponential based discriminant locality preserving projections to address the SSS problem of DLPP. By using the matrix exponential, the SSS problem is overcome elegantly and no discriminant information is lost for MEDLPP.
(2) We present an efficient procedure for solving MEDLPP, which, to our best knowledge, is the first efficient procedure for solving matrix exponential based method. Besides, the main idea for solving MEDLPP efficiently is also generalized to other matrix exponential based methods.

The remainder of the paper is organized as follows. In Section 2, we introduce some related works, i.e., LDA and DLPP. In Section 3, we propose the matrix exponential based discriminant locality preserving projections (MEDLPP) and the efficient procedure for solving MEDLPP. We report the experiment results in Section 4. Finally, we conclude the paper in Section 5.

2. Related works

2.1. Linear discriminant analysis (LDA)

Given a data matrix $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in \mathrm{R}^{d\times n}$, where $\mathbf{x}_i \in \mathrm{R}^d$, for $i = 1, \ldots, n$, is the ith training sample and assume that the data matrix X is grouped as $X = [X_1, \ldots, X_k]$, where $X_i \in \mathrm{R}^{d\times n_i}$ consists of the n_i data samples from the ith class and $\sum_{i=1}^{k} n_i = n$. Let N_i be the set of column indices that belong to the ith class, i.e., $\mathbf{x}_j$, for $j \in N_i$, belongs to the ith class. In the LDA method, there are three scatter matrices, which are called as within-class, between-class and total scatter matrices, respectively, and are defined as follows (Duda et al., 2000):

$$S_w = \sum_{i=1}^{k} \sum_{j\in N_i} (\mathbf{x}_j - \mathbf{m}_i)(\mathbf{x}_j - \mathbf{m}_i)^T \tag{1}$$

$$S_b = \sum_{i=1}^{k} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \tag{2}$$

$$S_t = \sum_{j=1}^{n} (\mathbf{x}_j - \mathbf{m})(\mathbf{x}_j - \mathbf{m})^T = S_b + S_w \tag{3}$$

where $\mathbf{m}_i = \frac{1}{n_i} X_i \mathbf{e}_i$ is the mean of the ith class, where $\mathbf{e}_i = (1, 1, \ldots, 1)^T \in \mathrm{R}^{n_i}$ and $\mathbf{m} = \frac{1}{n} X\mathbf{e}$ is the global mean of the dataset, where $\mathbf{e} = (1, 1, \ldots, 1)^T \in \mathrm{R}^n$.

The objective function of LDA is to find an optimal projection matrix G that maximizes the Fisher's criterion (Duda et al., 2000):

$$G = \arg\max_G tr\left(\left(G^T S_w G\right)^{-1} G^T S_b G\right) \tag{4}$$

where $tr(\cdot)$ is the trace operator. It is well-known that the optimal projection G can be obtained by solving the eigenvalue problem as follows:

$$S_b G = \lambda S_w G \tag{5}$$

or

$$S_w^{-1} S_b G = \lambda G \tag{6}$$

As shown in Eq. (6), LDA cannot work when S_w is singular. The PCA is usually used to overcome the singularity of S_w.