Neurocomputing 000 (2017) 1–14

Contents lists available at ScienceDirect

Neurocomputing

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

Exponential elastic preserving projections for facial expression recognition

Sen Yuan, Xia Mao*

School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

ARTICLE INFO

Article history:
Received 21 December 2016
Revised 10 May 2017
Accepted 29 August 2017
Available online xxx

Communicated by Xiaofeng Zhu

Keywords:
Facial expression recognition
Dimensionality reduction
Small sample size problem
Manifold learning
Matrix exponential

ABSTRACT

As a typical manifold learning method, elastic preserving projections (EPP) can well preserve the local geometry and the global information of the training set. However, EPP generally suffers from two issues: (1) the algorithm encounters the well known small sample size (SSS) problem; (2) the algorithm is based on the adjacent graph such that it is sensitive to the size of neighbors. To address these problems, we propose a novel method called exponential elastic preserving projections (EEPP), principally for facial expression recognition. By utilizing the properties of matrix exponential, EEPP is not only able to exploit the manifold structure of data, but also can get rid of the issues mentioned above. Experiments conducted on the synthesized data and several benchmark databases illustrate the effectiveness of our proposed algorithm.

1. Introduction

During the past couple of years, facial expression recogniton (FER) [1,2] has been extensively studied in the area of computer vision and pattern recognition due to its potential application in human-computer interaction and affective computing. A face image is usually represented as a data point in high-dimensional space. Often, such high-dimensional data lies close to a low-dimensional structure corresponding to a certain category to which the data belongs [3]. So dimensionality reduction (DR), which aims at exploring the low-dimensional subspace structure embedded in data, plays an important role in FER tasks. In many existing DR algorithms, principal components analysis (PCA) [4] and linear discriminant analysis (LDA) [5] are two classical linear DR methods. PCA attempts to project the data along an optimal direction by maximizing the variance matrix of data. Unlike PCA, LDA is a supervised method, which utilizes the label information and seeks to find a projection direction by maximizing the inter-class scatter and meanwhile minimizing the inner-class scatter. However, LDA usually suffers from the small sample size (SSS) problem. This stems from generalized eigen-problems with singular matrices. To overcome this limitation, many variants of LDA such as 2D-LDA [6], LDA/GSVD [7], LDA/QR [8], LDA/FKT [9], have been proposed in recent years. Moreover, to increase the degree of freedom and avoid the overfitting problem for 2D-LDA, Chang et al. [10] proposed compound rank-k projection (CRP), which utilizes multiple projection models to enhance the discriminant ability.

To our best knowledge, linear DR methods may fail to discover the underlying nonlinear structure, in which the high-dimensional image information in the real world lies. To remedy this deficiency, a large number of manifold learning algorithms have been proposed to discover the intrinsic nonlinear structure of the facial images. The most well-known manifold learning algorithms include locally linear embedding (LLE) [11], isometric feature mapping (ISOMAP) [12] and Laplacian eigenmaps (LE) [13]. LLE is an unsupervised method which focuses on local neighborhood of each data point and preserves the minimum linear reconstructing with neighborhood in the embedding space. ISOMAP determines the low-dimensional representations of original data by preserving geodesic distances between pairs of data points. LE is designed to preserve proximity relations of neighbor points by utilizing an undirected weighted graph. Unfortunately, these manifold learning methods usually suffer from the out-of-sample problem [14]. This is because they are defined only according to the training data, when new samples come, they cannot map them directly. To tackle this problem, He et al. proposed locality preserving projection (LPP) [15] and neighborhood-preserving embedding (NPE) [16]. In 2005, Pang et al. proposed neighborhood preserving projection (NPP) [17]. In order to reconstruct the data, Kokiopoulou and Saad proposed orthogonal neighborhood preserving projection (ONPP) [18], which forces the projection matrix to be orthogonal by solving the ordinary eigenvalue problem. To make sure that all the basis functions obtained by LPP are orthogonal, Cai et al.

* Corresponding author.
E-mail address: moukyou@buaa.edu.cn (X. Mao).

https://doi.org/10.1016/j.neucom.2017.08.067
0925-2312/

Please cite this article as: S. Yuan, X. Mao, Exponential elastic preserving projections for facial expression recognition, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.08.067

[19] also put forward orthogonal locality preserving projection (OLPP). The massive experiments [18–20] have shown that forcing an orthogonal relationship between the projection directions is more effective for preserving the manifold structure of high-dimensional data. Furthermore, in order to improve the discriminant power for classification tasks, lots of manifold-learning-based discriminant analysis algorithms [21–25] have been proposed. In 2014, Chang et al. [26] integrated manifold learning with clustering and proposed Spectral Shrunk Clustering (SSC), which has obtained quite promising clustering performance.

However, there exists a common problem with current manifold learning methods; that is, they only character the locality of samples such that they might not necessarily discover the most important manifold for pattern discrimination tasks. A reasonable approach should be one that integrates both nonlocal (global) and local structure into the objective functions of manifold learning. Yang et al. [27] proposed unsupervised discriminant projection (UDP). UDP introduces the concept of non-locality and can obtain the low-dimensional representation of data by maximizing the ratio of nonlocal scatter to local scatter. Zhang et al. [28] proposed complete global-local LDA (CGLDA) to incorporate three kinds of local information into LDA. In the literature [29], the authors proposed elastic preserving projections (EPP) which considers both the local structure and the global information of data. Luo et al. [30] added the discriminant information and orthogonal constraint into EPP, and proposed discriminant orthogonal elastic preserving projections (DOEPP). Recently, low-rank representation (LRR) [31–33], which is claimed to capture the global structure of the data, has received considerable interest. By solving the nuclear norm minimization, in general, LRR aims to find the lowest-rank representation among all the candidates that represent all vectors as the linear combination of the bases in a dictionary. Based on LRR, Liu et al. [34] proposed Laplacian regularized LRR (LapLRR), which discovers both the global Euclidean and local manifold structure of data, and is expected to have more discriminant power than LRR. In [35], a non-negative sparse hyper-Laplacian regularized LRR model (NSHLRR) was proposed. NSHLRR not only can represent the global low-dimensional structures, but also capture the local similarity information among data.

From the view of manifold learning, EPP aims to capture the local manifold structure and the global geometrical properties. However, in real world applications, EPP also has two main key issues. One is that the performance of EPP is sensitive to the size of neighbors. EPP constructs two graph models: the undirected neighborhood graph and the global graph. Usually, the most popular neighborhood graph construction manner is based on the K nearest neighbor criteria. Once the neighborhood graph is constructed, the edge weight is assigned by the heat kernel function. Unfortunately, such neighborhood graph needs to be artificially defined in advance, so it does not necessarily fit the intrinsic local structure of data. Still worse, the performance of EPP is seriously sensitive to the size of neighbors. The other one is the fact that it suffers from the small sample size problem (SSS). When the dimension of image sample is larger than the number of samples, the constraint matrix of EPP will be singular. To deal with this problem, a traditional approach is to use PCA for preprocessing. However, a potential problem is that the PCA step may discard some useful information for classification task.

Therefore, to alleviate the above issues, we propose a novel manifold learning algorithm called exponential elastic preserving projections (EEPP), motivated by the previous works [36,37]. Through introducing the matrix exponential, EEPP is more robust to the variation of the neighborhood size K. Moreover, the positive definite property of matrix exponential can also make EEPP avoid the SSS problem and obtain more valuable information.

The rest of this paper is organized as follows. We briefly review the LPP and EPP algorithm in Section 2. In Section 3, we introduce the proposed EEPP algorithm in detail. In Section 4, the experiments on two well-known synthetic manifold data sets and three widely used facial expression databases are carried out to demonstrate the effectiveness of the proposed method. Finally, Section 5 concludes the paper.

2. Review of the related work

Given a sample set $\mathbf{X} = \{x_1, x_2, \dots, x_N\} \in \mathbb{R}^{D\times N}$, the objective of dimensionality reduction is to find a transformation matrix $\boldsymbol{W}$ of size $D\times d$ to map: $y_i = \boldsymbol{W}^T x_i$, $y_i \in \mathbb{R}^{d\times N}$, where $d < D$, such that y_i is easier to be distinguished in the projective subspace.

2.1. Locality preserving projections

LPP [15] is an optimal linear approximation to LE. It only attempts to preserve the local structure of samples in the low-dimensional projective subspace. The local structure of data is measured by an adjacency graph G, constructed by K nearest neighbors. The corresponding similarity matrix $\mathbf{S}$ is defined by using the heat kernel function as:

$$S_{ij} = \begin{cases} \exp\left(-\|x_i - x_j\|_2^2/2t^2\right), & \text{if } x_j \in O(K, x_i) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $O(K,x_i)$ denotes the set of K nearest neighbors of x_i and t is a kernel. LPP defines the projected points in the form $y_i = \boldsymbol{W}^T x_i$. Therefore, its objective function can be obtained by solving the following minimization problem:

$$\begin{aligned} &\min \boldsymbol{W}^T \boldsymbol{XLX}^T \boldsymbol{W} \\ &\text{s.t. } \boldsymbol{W}^T \boldsymbol{XDX}^T \boldsymbol{W} = \mathrm{I}. \end{aligned} \tag{2}$$

where $\boldsymbol{L} = \boldsymbol{D}\text{-}\boldsymbol{S}$ is the Laplacian matrix. $\mathbf{D}$ is a diagonal matrix and its entries $\mathrm{D_{ii}} = \sum_j \mathrm{S_{ij}}$, $\mathrm{D_{ii}}$ measures the local density around x_i. The objective function incurs a heavy penalty if two neighbour points are mapped far apart. Thus, minimizing the function is an attempt to make sure that if x_i and x_j are close, then their corresponding projections y_i and y_j are close, as well. Finally, the optimization problem can be converted to the smallest eigenvalue of the following generalized eigen-problem:

$$\mathbf{XLX}^{\mathrm{T}}\mathrm{w} = \lambda \mathbf{XDX}^{\mathrm{T}}\mathrm{w} \tag{3}$$

Once the d eigenvectors are computed, the low-dimensional embedding results can be obtained by $y_i = \boldsymbol{W}^T x_i$, where $\boldsymbol{W} = [w_1, w_2, \dots, w_d]$.

2.2. Elastic preserving projections

Elastic preserving projections (EPP) [29] incorporates the advantage of both the local geometry and global information of data. The whole procedure of EPP can be decomposed into four steps:

1 *Local graph construction and selection of the weights*: The K nearest neighbor method is adopted to construct the local graph. If x_j is among the K nearest neighbors ofx_i, then put a direct edge from node i to node j. And the corresponding weights are defined as

$$S_{local}^{i,j} = \begin{cases} \exp\left(-\|\mathrm{x_i} - \mathrm{x_j}\|_2^2/2\mathrm{t}^2\right), & \text{if nodes i and j are connected} \\ 0, & \text{otherwise} \end{cases}$$

2 *Global graph construction and selection of the weights*: To preserve the global information and reflect the relationship between any two samples, the global weight matrix is defined as

$$S_{global}^{i,j} = \begin{cases} x_i - x_{j2}^{2} \exp\left(-x_i - x_{j2}^{2}/2\sigma^2\right), & i \neq j \\ 0, & i = j \end{cases}$$