Pattern Recognition 48 (2015) 1734–1742

Contents lists available at ScienceDirect

Pattern Recognition

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

Double adjacency graphs-based discriminant neighborhood embedding

Chuntao Ding [a], Li Zhang [a,b,*]

[a] School of Computer Science and Technology, Soochow University, Suzhou 215006, China
[b] Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, Suzhou 215006, China

ARTICLE INFO

Article history:
Received 8 January 2014
Received in revised form
3 July 2014
Accepted 30 August 2014
Available online 16 September 2014

Keywords:
Supervised learning
Discriminant neighborhood embedding
Manifold learning
Face recognition
Double adjacency graphs

ABSTRACT

Discriminant neighborhood embedding (DNE) is a typical graph-based dimensionality reduction method, and has been successfully applied to face recognition. By constructing an adjacency graph, aiming to keep the local structure for original data in the subspace, it is able to find the optimal discriminant direction effectively. Not for every sample does DNE set up a link between it and its heterogeneous samples when constructing the adjacency graph, which would result in a small between-class scatter. Motivated by this fact, we develop an extension of DNE, called double adjacency graphs-based discriminant neighborhood embedding (DAG-DNE) by introducing two adjacency graphs, or homogeneous and heterogeneous neighbor adjacency graphs. In DAG-DNE, neighbors belonging to the same class are compact while neighbors belonging to different classes become separable in the subspace. Thus, DAG-DNE could keep the local structure of a given data and find a good projection matrix for them. To investigate the performance of DAG-DNE, we compare it with the state-of-the-art dimensionality reduction techniques such as DNE and MFA on several publicly available datasets. Experimental results show the feasibility and effectiveness of the proposed method.

1. Introduction

Dimensionality reduction is an effective technique in machine learning and pattern recognition, which has been successfully applied to practical data preprocessing, such as face recognition [1,2] and image retrieval [3,4]. Dimensionality reduction techniques can not only reduce the computational complexity and avoid the curse of dimensionality problem, but also improve the classification performance. A general framework for dimensionality reduction was proposed in [5], which defines a general process according to describing different purposes of dimensionality reduction methods. Within this general framework, traditional dimensionality reduction techniques can be grouped into two categories: unsupervised and supervised ones. Unsupervised methods deal with data without labels, including principle component analysis (PCA) [6,7], locality preserving projection (LPP) [8,9], locally linear embedding (LLE) [10], neighborhood preserving embedding (NPE) [11] and so on. The goal of supervised methods is to obtain good classification performance and low complexity by projecting labeled data into a low-dimensional subspace. The typical supervised methods include linear discriminant analysis (LDA) [12], local discriminant embedding (LDP) [13], supervised optimal locality preserving projection (SoLPP) [14], marginal Fisher analysis (MFA) [12,5], discriminant neighborhood embedding (DNE) [15] and others.

As a classical unsupervised algorithm, PCA projects the original data into the low-dimensional subspace which could contain the global information of original data. However, PCA is not very effective in dealing with manifold data. LLE proposed in [10] first uses linear coefficients, which can reconstruct a given sample by its neighbors to represent the local geometry, and then seeks a low-dimensional embedding where these coefficients are still suitable for reconstruction. However, mappings yielded by LLE are only defined on the training data. It remains unclear how to naturally evaluate the mappings on novel testing points. In other words, LLE suffers the problem of out-of-sample. As a remedy, NPE [11], orthogonal neighborhood preserving projection (ONPP) [16] and LPP [8,9] were proposed successively. Both NPE and ONPP which are linear approximations to LLE aim at preserving the local manifold structure and can also solve the out-of-sample problem of LLE. The main difference between NPE and ONPP is that they solve generalized eigenvalue problems with different constraints. LPP finds an embedding to preserve local information and can be simply extended to any new samples. Since LPP is unsupervised, it cannot work well in classification tasks.

LDA can find effective project directions for classification tasks by simultaneously maximizing the between-class scatter and

* Corresponding author.

http://dx.doi.org/10.1016/j.patcog.2014.08.025
0031-3203/© 2014 Published by Elsevier Ltd.

minimizing the within-class scatter. Although being a good choice for dimensionality reduction, LDA still has several problems. For example, LDA requires the assumption that all samples contribute equivalently for discriminative dimensionality reduction and only considers the global Euclidean structure. In order to overcome these problems, marginal Fisher analysis (MFA) has been proposed in [12,5]. As an extension of LDA, MFA can effectively solve small sample problems by constructing the within-class adjacency graph and the inter-class adjacency graph to keep the local structure for samples. Similar to LDA, MFA can also find the projection directions that maximize the inter-class scatter and minimize the intra-class scatter simultaneously. As another supervised method, DNE proposed in [15] could find the best projection directions by using spectrum analysis. In DNE, an adjacency graph is constructed to keep local structure by distinguishing between homogeneous and heterogeneous neighbors. However, when constructing the adjacency graph, not for each point does DNE construct a link between it and its heterogeneous points. In doing so, DNE would obtain a small between-class scatter in the subspace.

To remedy this, this paper presents a new supervised discriminant subspace learning algorithm, called double adjacency graphs-based discriminant neighborhood embedding (DAG-DNE). In DAG-DNE, each sample is respectively linked to its homogeneous and heterogeneous neighbors by constructing double adjacency graphs. The idea of constructing double adjacency graphs comes from MFA. As a consequence, balanced links are produced, which solves the problem existed in DNE. In DAG-DNE, neighbors belonging to the same class are compact while neighbors belonging to different classes become separable in the subspace. Thus, DAG-DNE could keep the local structure of a given data and find a good projection matrix for them.

The remainder of this paper is organized as follows. Section 2 introduces the related works on MFA and DNE. In Section 3, we propose DAG-DNE and discuss its connections to MFA and DNE. Section 4 reports simulation experimental results and Section 5 concludes this paper.

2. Related works

In this section, we briefly review MFA and DNE, which are the foundation of our work. Both MFA and DNE reduce the dimensionality of high-dimensional data and keep the local structure via graph embedding. Let $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ be a set of training samples, where $\mathbf{x}_i \in \mathbb{R}^d, y_i \in \{1, 2, \ldots, c\}$ is the class label of $\mathbf{x}_i$, d is the dimensionality of samples, N is the number of training samples, and c is the number of classes. We try to learn a linear transformation that can project the data in the original d-dimensional space to a new r-dimensional subspace in which the samples are denoted as $\{(\mathbf{v}_i, y_i)\}_{i=1}^N$ with $r \ll d$, and

$$\mathbf{v}_i = \mathbf{P}^T \mathbf{x}_i \tag{1}$$

where $\mathbf{P} \in R^{d \times r}$ is the projection matrix. Both MFA and DNE try to find an optimal projection matrix which is the best for classification tasks.

2.1. Marginal Fisher analysis

MFA proposed based on LDA can overcome some shortcomings of LDA. Since real-world data do not always satisfy a Gaussian distribution, LDA could not work well on it. However, MFA could project the data into a discriminative feature space in which the data is clearly separated.

In MFA, there are two graphs, or intra-class compactness and inter-class separability graphs. In the intra-class compactness graph, for each sample $\mathbf{x}_i$, we set the intra-class adjacency matrix $W_{ij}^w = W_{ji}^w = 1$ if $\mathbf{x}_j$ is among the k-nearest neighbors of $\mathbf{x}_i$ in the same class, otherwise $W_{ij}^w = W_{ji}^w = 0$. In the inter-class graph, for each sample $\mathbf{x}_i$, we set the inter-class adjacency matrix $W_{ij}^b = W_{ji}^b = 1$ if $\mathbf{x}_j$ is among the k-nearest neighbors of $\mathbf{x}_i$ in the different classes, otherwise $W_{ij}^b = W_{ji}^b = 0$.

MFA aims at finding the optimal projection directions via the following Marginal Fisher Criterion:

$$\min_{\mathbf{P}} \frac{\mathrm{tr}(\mathbf{P}^T \mathbf{X}(\mathbf{D}^w - \mathbf{W}^w)\mathbf{X}^T \mathbf{P})}{\mathrm{tr}(\mathbf{P}^T \mathbf{X}(\mathbf{D}^b - \mathbf{W}^b)\mathbf{X}^T \mathbf{P})} \tag{2}$$

where tr() is the trace of a matrix, $D_{ii}^w = \sum_j W_{ij}^w$, $D_{ii}^b = \sum_j W_{ij}^b$, and $\mathbf{P}$ is composed of the optimal r projection vectors.

2.2. Discriminant neighborhood embedding

DNE is a supervised subspace learning method, which forms a compact submanifold for data in the same class in the embedded low-dimensional subspace. Simultaneously, DNE tries to make the gaps among submanifolds for different classes as wide as possible. First, DNE is to construct the adjacency graph by using k -nearest neighbors. The adjacency weight matrix $\mathbf{W}$ is defined as

$$W_{ij} = \begin{cases} +1, & \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors and } y_i = y_j \\ -1, & \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors and } y_i \neq y_j \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Second, DNE is to solve the following optimization problem:

$$\begin{cases} \min \sum_{i,j} ||\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \mathbf{x}_j||^2 W_{ij} \\ \text{s.t. } \mathbf{P}^T \mathbf{P} = \mathbf{I} \end{cases} \tag{4}$$

where $\mathbf{I}$ is the identify matrix. The objective function in (4) can be further derived as follows:

$$\begin{aligned} \sum_{i,j} ||\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \mathbf{x}_j||^2 W_{ij} &= 2\sum_{i,j} (\mathbf{x}_i^T \mathbf{P}\mathbf{P}^T \mathbf{x}_i - \mathbf{x}_i^T \mathbf{P}\mathbf{P}^T \mathbf{x}_j) W_{ij} \\ &= 2\,\mathrm{tr}\{\mathbf{P}^T \mathbf{X}(\mathbf{D} - \mathbf{W})\mathbf{X}^T \mathbf{P}\} \\ &= 2\,\mathrm{tr}(\mathbf{P}^T \mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{P}) \end{aligned} \tag{5}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ and $D_{ii} = \sum_j W_{ij}$. Thus, the optimization problem (4) can be rewritten as

$$\begin{cases} \min_{\mathbf{P}} \mathrm{tr}(\mathbf{P}^T \mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{P}) \\ \text{s.t. } \mathbf{P}^T \mathbf{P} = \mathbf{I} \end{cases} \tag{6}$$

The projection matrix $\mathbf{P}$ can be found by solving the generalized eigenvalue problem as follows:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{P} = \lambda \mathbf{P} \tag{7}$$

Thus $\mathbf{P}$ is composed of the optimal r projection vectors corresponding to the r smallest eigenvalues.

3. Double adjacency graphs-based discriminant neighborhood embedding

In this section, we develop a novel supervised subspace learning method called double adjacency graphs-based discriminant neighborhood embedding (DAG-DNE). In DNE, the adjacency links between a sample and its heterogeneous neighbors are much weaker than those between the sample and its homogeneous neighbors. When considering the adjacency relationship between samples, DAG-DNE establishes two adjacency graphs for a sample, one for its homogenous neighbors, and the other for its heterogeneous neighbors. As a result, balanced links are produced.