Pattern Recognition 57 (2016) 179–189

Contents lists available at ScienceDirect

Pattern Recognition

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

# Semi-supervised linear discriminant analysis for dimension reduction and classification

Sheng Wang [a], Jianfeng Lu [a,*], Xingjian Gu [a], Haishun Du [b], Jingyu Yang [a]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Institute of Image Processing and Pattern Recognition, Henan University, Kaifeng 475004, China

ARTICLE INFO

*Article history:*
Received 30 July 2015
Received in revised form
12 November 2015
Accepted 24 February 2016
Available online 8 March 2016

*Keywords:*
Dimension reduction
Semi-supervised learning
Linear discriminant analysis
Data classification

ABSTRACT

When facing high dimensional data, dimension reduction is necessary before classification. Among dimension reduction methods, linear discriminant analysis (LDA) is a popular one that has been widely used. LDA aims to maximize the ratio of the between-class scatter and total data scatter in projected space, and the label of each data is necessary. However, in real applications, the labeled data are scarce and unlabeled data are in large quantity, so LDA is hard to be used under such case. In this paper, we propose a novel method named semi-supervised linear discriminant analysis (SLDA), which can use limited number of labeled data and a quantity of the unlabeled ones for training so that LDA can accommodate to the situation of a few labeled data available. Assuming that **F** represents the calculated class indicator matrix of the training data and **Y** denotes the true label of the labeled data, the objective function contains two parts: one is the criterion of LDA (which is a function of projection **W**, and a class indicator matrix **F**), the other is the difference between the true data label and calculated label of these labeled data. As far as we know, there is no closed-form solution to the objective function. To solve such problem, we develop an iterative algorithm which calculates the class indicator matrix and the projection alternatively. The convergence of the proposed iterative algorithm is proved and confirmed by experiments. The experimental results on eight datasets show that the performance of SLDA is superior to that of traditional LDA and some state-of-the-art semi-supervised algorithms.


## 1. Introduction

In the scenario of machine learning, we are often involved in dealing with high-dimensional data, which needs more storage and computational time. There are lots of approaches to address this problem, and dimension reduction is the simplest one. Up to now, lots of works have been devoted to dimension reduction and successfully applied to image classification [1–6]. These methods are broadly categorized into two classes: unsupervised methods and supervised methods.

Due to good performance of supervised methods, lots of methods on supervised dimension reduction have been proposed, such as: linear discriminant analysis (LDA) [1], locality preserving projection (LPP) [7], neighborhood preserving embedding (NPE) [8], marginal Fisher analysis (MFA) [2], local Fisher discriminant analysis (LFDA) [9], and adaptive slow feature discriminant analysis [6]. Among all these algorithms, linear discriminant analysis is the most commonly used dimension reduction method for classification. It aims to maximize the distance between the projected data of different classes and minimizes the distance between the projected data of the same classes.

As we all know, these supervised dimension reduction methods need the label of each sample during the training stage. However, in the scenario of real applications, the data may not have label due to various reasons, e.g. it is expensive and time consuming to collect sufficient labeled data [10], or label is missing. Besides, these supervised dimension reduction algorithms cannot utilize unlabeled data. Further, they cannot totally exploit the discriminant information when label information of the training samples is not sufficient. To cope with these problems, semi-supervised learning algorithms are proposed [11–17], such as: semi-supervised discriminant analysis [18], flexible manifold embedding [11], semi-supervised dimensionality reduction (SSDR) [12] and semi-supervised discriminant analysis(SDA) [13]. Roughly, most of these algorithms consist of two stages: (1) model the geometric relationship (including local geometry or global geometry of training data) between all data points; (2) incorporate the geometric relationships to the objective function of the classic algorithms. In Ref. [12], Zhang et al. firstly modeled the geometric relationships using the global scatter of the total training

DOI of original article: http://dx.doi.org/10.1016/j.bbagen.2015.12.016
* Corresponding author. Tel.: +86 25 84315751x824.
*E-mail addresses:* wangsheng1910@163.com (S. Wang), lujf@njust.edu.cn (J. Lu), guxingjian163@163.com (X. Gu).

http://dx.doi.org/10.1016/j.patcog.2016.02.019
0031-3203/

data, then they incorporated global scatter into the framework of maximum margin criterion (MMC) and proposed semi-supervised dimensionality reduction (SSDR). As we all know, the adjacent graph representation can provide the local geometry of the data. In order to preserve the local geometry of the data, Cai et al. [13] proposed semi-supervised discriminant analysis(SDA) which maximizes the separability between different classes (which utilizes the labeled samples) and preserves the local geometry of the data using the adjacent graph. In Refs. [19,11], Nie et al. incorporated a term related to manifold smoothness into multidimensional least squares linear discriminant analysis (MLSLDA) [20], and proposed flexible manifold embedding (FME) and Laplacian regularized least squares (LapRLS).

As study continues, more and more works [21–23] devoted to semi-supervised learning for dimension reduction. They are different from semi-supervised learning methods mentioned in last paragraph, which divided the stage (2) into two steps: (a) get the soft label (not true label) of unlabeled data using the labeled samples and the geometric relationships learned by stage (1); (b) utilize the soft label of unlabeled data and true label of labeled data to construct the objective function. In Ref. [21], Nie et al. proposed a method named semi-supervised orthogonal discriminant analysis via label propagation (SODA). In SODA, they firstly calculated the labels of these unlabeled data and computed the projection matrix by maximizing the objective function of orthogonal discriminant analysis. Lu et al. [22] proposed cost-sensitive semi-supervised discriminant analysis ($CS^3DA$) which also computed the virtual labels of the unlabeled data and utilized labels of the total training samples. Zhang and Yeung [23] proposed semi-supervised generalized discriminant analysis (SSGDA) which computed labels of unlabeled data by constrained concave–convex procedure and unlabeled data then be used to augment the original labeled data for performing generalized discriminant analysis (GDA). Labels of unlabeled data used in these methods are calculated using the adjacent graph or the scatter of the data. However, both the adjacent graph and the scatter are constructed using the original data, so they may have one or two of the following limitations: (1) they may ignore the relation among classes during computing the labels of the unlabeled data; (2) they cannot guarantee good performance during computing labels because original data may be contaminated by noise, which is brought by various facial expressions, illumination in face recognition.

From the above analysis, we know that the original feature may be contaminated by noise or contain useless information. Then, calculated labels based on the original feature or geometric relationships of training samples will be quite different from their true labels. Therefore, mapping using these calculated labels may not be optimal for classification. Further, we should consider the relation among classes during computing the labels of the unlabeled data. Considering above two issues, in this paper, we propose a new method named semi-supervised linear discriminant analysis (SLDA). The label matrix **F** (**F** represents labels of the samples by 1-of-k scheme) is calculated using mapped data by optimizing the criterion of SLDA which has more discriminant information and less noise. Then, the calculated labels will be employed to construct the between-class scatter matrix and compute the mappings by eigen decomposition, and SLDA can be implemented by iterating the following two steps: (1) computes projection; (2) calculates the label matrix **F** using the projected data. These two steps are executed alternatively before the objective function converges. From the definition of label matrix $\mathbf{F}_{ij}\geq 0, \mathbf{F}_{ij}\in\left\{0,\frac{1}{\sqrt{n_j}}\right\}, \sum_j F.j=\sqrt{n_j}$ and $\mathbf{F}^T F=I$, we know that optimizing the objective functions of **F** (given **W**) is a typical NP-hard problem. So, we should relax the elements of class indicator matrix **F** from discrete values to continuous ones. To get a more accurate solution, we constrain **F** to be non-negative and orthogonal between columns. The objective function has two variables and no closed-form solution. To address such optimization, we develop an iterative procedure. In Sections 3 and 4, we will prove the convergence of proposed optimized method and experimental results also manifest it. The experimental results on 8 datasets show that our method outperforms most of state-of-the-art semi-supervised dimension reduction methods for classification.

This paper is organized as follows: we will analyze the classic LDA in Section 2. Then, we will propose an algorithm named semi-supervised linear discriminant analysis for dimension reduction in Section 3. In Section 4, the performance of SLDA and some classic semi-supervised dimension reduction methods will be compared. At last, some conclusions will be concluded in Section 5.

## 2. Linear discriminant analysis

In this section, we will give a brief overview of classical LDA. Given a dataset $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_n]\in R^{D\times n}$, where $\{\mathbf{x_i}\}_{i=1}^n$ represents the labeled data. **X** can be partitioned into $C$ disjoint classes and its mean is a zero vector. $y_i(y_i\in\{1,2,\ldots,C\})$ denotes the label of $\mathbf{x}_i(i=1,\ldots,n)$. Without loss of generality, we use 1-of-K scheme to represent the class membership of samples such as, $\mathbf{Y}_{ij}=1$ denotes $y_i=j$; otherwise, $\mathbf{Y}_{ij}=0$. Further, **Y** can be scaled by $\mathbf{Y}=\mathbf{Y}(\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{\frac{1}{2}}$. Classic LDA computes a linear projection that maps $\mathbf{x}_i$ into a low dimensional space: $\mathbf{W}:\tilde{x}_i=W^Tx_i$. In LDA, three scatter matrices, named within-class, between-class and total scatter matrices are defined as:

$$\mathbf{S}_w=\frac{1}{n}\sum_{i=1}^{C}\sum_{y_j=i}(\mathbf{x}_j-\mathbf{m}_i)(\mathbf{x}_j-\mathbf{m}_i)^T$$
$$\mathbf{m}_i=\frac{1}{n_i}\sum_{y_j=i}\mathbf{x}_j \tag{1}$$

$$\mathbf{S}_b=\sum_{i=1}^{C}\frac{n_i}{n}(\mathbf{m}_i-\mathbf{m})(\mathbf{m}_i-\mathbf{m})^T \tag{2}$$

$$\mathbf{S}_t=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\mathbf{m})(\mathbf{x}_i-\mathbf{m})^T \tag{3}$$

where $n_i$ denotes the number of samples in class $i$.

From the above definition, we know that $\mathbf{S_t}=S_b+S_w$. If data have been centered by subtracting their mean, between-class and total scatter matrices can be constructed by:

$$\mathbf{S}_b=\sum_{i=1}^{C}\frac{n_i}{n}\mathbf{m}_i\mathbf{m}_i^T=\frac{1}{n}\mathbf{X}\mathbf{Y}\mathbf{Y}^T\mathbf{X}^T \tag{4}$$

$$\mathbf{S}_t=\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T=\frac{1}{n}\mathbf{X}\mathbf{X}^T \tag{5}$$

The optimal transformation of LDA is computed by solving the following problem:

$$\mathbf{W}=\max_{\mathbf{W}}\frac{\mathrm{tr}\left\{\mathbf{W}^T\mathbf{S}_b\mathbf{W}\right\}}{\mathrm{tr}\left\{\mathbf{W}^T\mathbf{S}_t\mathbf{W}\right\}} \tag{6}$$

## 3. Semi-supervised linear discriminant analysis

In this section, we will detail the proposed method semi-supervised linear discriminant analysis. The training data is defined as $\mathbf{X}=[x_1,\ldots,x_m]\in R^{D\times m}$, where $\{\mathbf{x}_i\}_{i=1}^n(n<m)$ represents the labeled data and $\{\mathbf{x}_i\}_{i=n+1}^m$ represents the unlabeled data. **X**