Contents lists available at ScienceDirect

Neurocomputing

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

Recursively global and local discriminant analysis for semi-supervised and unsupervised dimension reduction with image analysis

Shangbing Gao [a,b,*], Jun Zhou [a], Yunyang Yan [a], Qiao Lin Ye [c]

[a] Faculty of Computer and Software Engineering, Jiangsu Provincial Key Laboratory for Advanced Manufacturing Technology, Huaiyin Institute of Technology, Huai'an, PR China
[b] Key Laboratory of Image and Video Understanding for Social Safety (Nanjing University of Science and Technology), Nanjing, PR China
[c] School of Information Technology, Nanjing Forestry University, Nanjing, PR China

ARTICLE INFO

Article history:
Received 6 April 2016
Received in revised form
22 July 2016
Accepted 8 August 2016
Communicated by: Dacheng Tao

Keywords:
Semisupervised dimension reduction
"Concave-convex" programming problem
Feature extraction
Fisher linear discriminant analysis
Manifold regularization

ABSTRACT

Semi-supervised discriminant analysis (SDA) is a recently-developed semi-supervised dimension reduction method for improving the performance of Fisher linear discriminant analysis (FDA), which attempts to mine the local structures of both labeled and unlabeled data. In this paper, we develop new semi-supervised and unsupervised discriminant analysis techniques. Our semi-supervised method, referred as to as recursively global and local discriminant analysis (RGLDA), is modeled based on the characterizations of "locality" and "non-locality", such that the manifold regularization in the formulation has a more direct connection to classification. The objective of RGLDA is a "concave-convex" programming problem based on the hinge loss. Its solution follows from solving multiple related SVM-type problems. In addition, we also propose a simple version (called URGLDA) for unsupervised dimension reduction. The experiments tried out on several image databases show the effectiveness of RGLDA and URGLDA.

1. Introduction

In many real applications, such as face recognition, we are usually faced with high-dimensional data. The data may be represented by original images or many kinds of visual features [45]. In such cases, exacting good features is crucial for mitigating the so-called "curse of dimensionality" and improving the performance of any pattern classifier. Dimensionality reduction techniques, such as Principal Component Analysis (PCA) [1,18] and Fisher Linear Discriminant Analysis (FDA) [2,14,35], are developed for this purpose. PCA and FDA have been widely applied in the fields of pattern recognition and computer vision. Many experimental studies have shown that FDA outperforms PCA significantly [3,4,36]. PCA can also be used for the data reconstruction. Same as PCA, the k-dimensional coding schemes [41,42] attempt to represent data using a set of representative k-dimensional vectors. Bian et al. [44] presented analysis on the bound of FDA.

Recently, there is much interest in developing manifold learning algorithms, e.g., Isometric Feature Mapping (ISOMAP) [5], Local Linear Embedding (LLE) [6], Laplacian Eigenmap (LE) [7], and Locality Preserving Projections (LPP) [8]. A main problem of ISOMAP, LLE and LE is that they y cannot map the unknown points [8]. LPP was proposed to deal with this problem [37,38]. Such an algorithm, however, have no direct connection to classification [9], which only characterizes local scatter and ignores the characterization of "non-locality". Recently, Yang et al. [9] proposed Unsupervised Discriminant Projection (UDP) to resolve this issue. UDP is modeled based on "locality" and "non-locality" [9]. In these unsupervised methods, the prior class information is not used.

Supervised learning algorithms may not generalize well enough when there is no sufficiently supervised information, although they generally outperform unsupervised learning algorithms [39,40]. Furthermore, collecting labeled data is generally more involved than collecting unlabeled clearly [10], since it requires expensive human labor, and a large amount of noise and outlier data are easy to be introduced. In [43], Liu et al. have attempted to use importance reweighting to solve the classification problems where the labeled samples are corrupted. In order to sufficiently exploit unlabeled data for better classification, many semi-supervised learning algorithms, such as Transductive SVM (TSVM) [11] and graph-based semi-supervised learning algorithms [10,12], have been recently developed. Among them, Manifold Regularization (MR) [10] is the most attractive [10]. However, these algorithms are developed for classification problems. Based on distance learning [46], Yu et al. [47] proposed semisupervised multiview distance metric learning. In dimension reduction, Cai et al., [13] proposed a

* Corresponding author.
E-mail address: luxiaofen_2002@126.com (S. Gao).

http://dx.doi.org/10.1016/j.neucom.2016.08.018
0925-2312/

novel method, called semi-supervised discriminant analysis (SDA). It not only preserves the discriminating information of labeled data but also the local structure of both labeled and unlabeled data. Specifically, both labeled and unlabeled data are used to build a graph Laplacian [13] that is incorporated into the FDA to smooth the mapping. SDA inherits the respective superiorities of FDA and manifold learning. However, SDA focuses only on the local geometry, such that it has no direct connection to classification [39]. A basic assumption behind SDA is that nearby points will have similar embeddings [13], which is consistent with the manifold assumption [10]. In practice, if the manifold assumption holds, the points from different classes could lie in different manifolds, which means that if we would like to obtain a better projection than that yielded by SDA, then, like UDP [9], the "non-locality" should be also taken into account in the framework of SDA, such that as much discriminant information of labeled and unlabeled points as possible can be mined. Based on SDA, Huang et al. [32] proposed TR-FLDA. By employing the similar idea in TR-FLDA, Dornaika et al. [33] proposed a new graph-based semisupervised DR (GSDR) method. However, these two methods suffer from the same problems of SDA. For existing eigenvalue-based dimension reduction techniques, it is difficult to introduce the sparsity in the projection matrix [15]. The last few years has seen few elegant sparse dimension reduction techniques in multi-class setting, such as Sparse PCA (SPCA) [15], which uses L_1-penalized regression on regular projection axes. With the similar technique, we can also develop a multi-class sparse SDA method. However, they are two-stage approaches. Intuitively, it is too naive to believe that there is no any information loss in the two-stage processing. We believe that sparsity should be introduced into the model by entirely following the genuine geometrical interpretation of SDA for the guarantee of obtaining a good performance.

In this paper, we propose a new semi-supervised dimension reduction method, called Recursively Global and Local Discriminant Analysis (RGLDA) which includes two steps. First, a novel dimension reduction framework for semi-supervised learning, called Global and Local Discriminant Analysis (GLDA), is proposed. In addition to incorporating the basic idea behind SDA of finding an optimal projection and estimating the local geometry of a relatively limited amount of labeled data as well as considerable unlabeled data, GLDA simultaneously mines the underlying non-local structure of labeled and unlabeled data. Second, the projections are generated by using the similar recursive procedure as suggested in RFDA [19]. The new algorithm is essentially developed from the SDA but has a significant performance advantage. Different from SDA which maximizes the squared sum of all the pairwise distances between class means, our method maximizes every pairwise distance between class means and allows some pairwise distances commit the maximization limit. Moreover, the prior work [16,17,32] on various recognition tasks has demonstrated that casting an eigenvalue-based problem as a related SVM-type problem can lead to better recognition rates. In addition, we extend RGLDA to a simple version (called RGLDA) for unsupervised dimension reduction. The main contributions of this study are summarized as follows:

1) We propose a framework for semi-supervised and unsupervised dimension reduction, whose unique idea can be used in most of the up-to-date semisupervised dimension reduction methods, such as TR-FLDA [33], and GSDR [34].
2) Our work characterizes not only the local but also the non-local quantities, such that the mapping has a direct connection to classification.
3) Our proposed framework is flexible. To be specific, the special formulation allows us to easily develop the sparse semi-supervised model by incorporating various regularization techniques into the formulation in future work. Unlike the recently proposed two-stage sparse methods, the sparsity is directly introduced to the model.

We organize this paper as follows. Section II gives a brief review of FDA and SDA. In section III, we give a basic framework for dimension reduction that casts an eigenvalue-based problem as an SVM-type problem, proposes GLDA, and shows its solution. Section IV develops recursive GLDA (RGLDA), which aims at generating multiple projection axes. In section V, a simplified dimension reduction method for unsupervised learning is proposed. In section VI, we evaluate our algorithms on several image databases. Section VII gives a method of constructing sparse dimension reduction methods based on the formulation of RGLDA. In the last section, we draw some conclusions.

2. Review of FDA and SDA

Let $\mathbf{x} = \{\mathbf{x}_i|\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l, \mathbf{x}_{l+1}, \mathbf{x}_{l+2} \ldots \mathbf{x}_m\} \in \mathbb{R}^{n\times m}$ be the data matrix, where m and n are the number and dimensionality of data, and $\mathbf{x}_i|_{i=1}^{l}$ and $\mathbf{x}_i|_{i=l+1}^{m}$ are labeled and unlabeled data, respectively. The labeled data is from c different classes. There are N_u samples in each class, which are represented by $\boldsymbol{D}_u$, $u = 1, \ 2, \ldots, c$. Define by $\mathbf{z} = \in \mathbb{R}^{d}(1 \le d \le n)$ a low-dimensional representation of a high-dimensional sample $\mathbf{x}$ in the original space, where d is the dimensionality of the reduced space. The purpose of dimension reduction is to seek for a transformation matrix $\mathbf{W}$, such that a lower representation $\mathbf{z}$ of the sample $\mathbf{x}$ can be calculated as $\mathbf{z} = \mathbf{W}^{\mathrm{T}}\mathbf{x}$, where " T" denotes the transpose. Other notations are listed in Table 1, each being explained when it is first used.

2.1. Fisher linear discriminant analysis (FDA)

FDA, as a popular supervised dimension reduction method, aims at finding the projection that simultaneously maximizes the between-class scatter and minimize the within-class scatter. That is, we can obtain the projection by maximizing the following objection function

$$J_F(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_B\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_W\mathbf{w}}, \tag{1}$$

where the within-class scatter matrix $\mathbf{S}_W$ is defined by

$$\mathbf{S_W} = \sum_{u=1}^{c}\sum_{x\in \boldsymbol{D}_u}(\mathbf{x}-\boldsymbol{\mu}^{(u)})(\mathbf{x}-\boldsymbol{\mu}^{(u)})^{\mathrm{T}} = \mathbf{X}^T\mathbf{L}_F\mathbf{X} \tag{2}$$

Table 1
Notations.

$\mathbf{S}_w \in \mathbf{R}^{n\times n}$	The within-class scatter;
$\mathbf{S}_B \in \mathbf{R}^{n\times n}$	The between-class scatter;
$\mathbf{S}_t \in \mathbf{R}^{n\times n}$	The global scatter;
$\boldsymbol{\mu}^{(u)} \in \mathbf{R}^{n}$	The mean vector of the samples in class u;
$\boldsymbol{\mu} \in \mathbf{R}^{n}$	The Total mean of the samples;
$\mathbf{H} \in \mathbf{R}^{m\times m}$	The adjacency matrix;
ρ	A regularization parameter;
v	A regularization parameter;
$\mathbf{D} \in \mathbf{R}^{m\times m}$	A diagonal matrix whose entries on diagonal are column or row sum of $\mathbf{H}$;
$\mathbf{L} = \mathbf{D} - \mathbf{H} \in \mathbf{R}^{m\times m}$	The Laplacian matrix;
$\mathbf{S}_L = \mathbf{XLX}^T \in \mathbf{R}^{n\times n}$	A graph scatter matrix;
f_i and g_i	Two real-values convex functions on a vector space $\mathbf{Z}$;
$T_1\{f, \mathbf{z}\}(\mathbf{z}')$	The first order Taylor expansion of f at location $\mathbf{z}$;
$\partial_{\mathbf{z}} f(\mathbf{z})$	The gradient of the function f at location $\mathbf{z}$;
$\xi \in \mathbf{R}^{m}$	The Hinge loss function;
$\mathbf{X}^{(u)}$	The sample matrix of class u;
$\mathbf{e}^{(u)}$	A column vector of ones of size($\mathbf{X}^{(u)}$) dimensions.