Neurocomputing 000 (2017) 1–15

Contents lists available at ScienceDirect

Neurocomputing

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

An iterative paradigm of joint feature extraction and labeling for semi-supervised discriminant analysis

Shougang Ren[a,b,d], Xingjian Gu[a,d,*], Peisen Yuan[a], Huanliang Xu[a,c]

[a] College of Information Science and Technology, Nanjing Agricultural University, Nanjing, 210095, China
[b] National Engineering and Technology Center for Information Agriculture, Nanjing, 210095, China
[c] Collaborative Innovation Center of Meat Production and Processing, Nanjing, 210095, China
[d] Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University), Fuzhou, 350121, China

ARTICLE INFO

Article history:
Received 23 February 2016
Revised 3 May 2017
Accepted 2 August 2017
Available online xxx

Communicated by Dr. Qi Li

Keywords:
Feature extraction
Semi-supervised learning
Uncorrelated constrain
Nonnegative constraint

ABSTRACT

Semi-supervised discriminant analysis (SDA) has been attracting much attention as it utilizes both of labeled and unlabeled data simultaneously. Current SDA method estimates the labels of unlabeled data just in the original data space. However, in real application, the original features are often contaminated and those original features contain much redundant information. Furthermore, accurate estimation for labels of unlabeled samples plays a critical role for semi-supervised discriminant analysis. In this paper, a novel method called Iterative Semi-supervised discriminant analysis (Iterative-SDA) is proposed, which takes an iterative manner to estimate the labels of unlabeled samples. Each iteration of Iterative-SDA mainly contains two steps: (1) Extract low-dimensional features based on the estimated labels F of samples; (2) Estimate the labels based on the extracted low-dimensional features. It is an NP hard problem to optimize the binary label matrix F, so we explicitly impose nonnegative and orthogonal constrains to relax binary label matrix F during the process of iteration. Experiments are done on four benchmark databases including AR, ORL, CMU PIE and COIL20, the experimental results demonstrate that the proposed method is more effective than LDA and some other state-of-art semi-supervised methods.

1. Introduction

Feature extraction is a fundamental and challenging issue in many areas such as computer vision, pattern recognition and gene expression array analysis. The goal of feature extraction is to seek for an optimal low-dimensional subspace, which can provide a compact representation of high-dimensional data. During the past decades, lots of feature extraction methods have been developed for clustering or classification tasks. Those methods can be broadly categorized into three parts including unsupervised methods, supervised methods and semi-supervised.

Principal Component Analysis (PCA) [1] is one of the most popular unsupervised feature extraction methods. It reduces the dimension of data while preserving the most representative information for optimal reconstruction. However, PCA has the inability to discover the nonlinear manifold structure. To address this problem, many manifold-based learning for dimensionality reduction methods are developed. The representative manifold-based learning methods include Isomap [2], Locally Linear Embedding (LLE)[3], and Laplacian Eigenmap (LE) [4]. Isomap preserves pairwise geodesic distance of observations in embedding space. LLE focuses on local neighborhood of each data point and preserves the minimal error of linear reconstructing with neighborhood in the low-dimensional space. LE is developed on Laplace Betltrami operator to preserve proximity relationship of data points. In order to overcome out-of-sample problem, He et al. [5] propose Neighborhood Preserving Embedding (NPE) to seek a linear subspace that preserves local structure based on the same principle of LLE. Locality Preserving Projections (LPP) [6,7] seeks a linear subspace to approximate nonlinear Laplacian Eigenmap. The aforementioned methods are unsupervised, which do not need to use the prior label information. During the past decades, lots of supervised methods have been proposed, such as linear discriminant analysis (LDA) [8], Margin Fisher Analysis (MFA) [9], Local Discriminant Embedding (LDE) [10], Maximum Margin Criterion (MMC) [11] and so on. Those discriminant methods aim to find a set of low-dimensional space to maximize between-class dissimilarity and minimize within-class dissimilarity simultaneously.

However, in many real word applications, the number of labeled data is often insufficient. A large quantity of data lacks of labels due to various reasons, e.g. it is labor expensive and time

* Corresponding author.
E-mail address: guxingjian163@163.com (X. Gu).

http://dx.doi.org/10.1016/j.neucom.2017.08.003
0925-2312/© 2017 Published by Elsevier B.V.

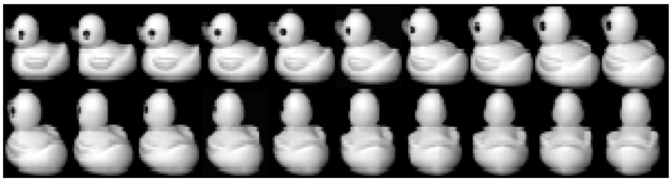

Fig. 1. Sample images in COIL20 database.

Table 1
The maximal average recognition accuracy (Mean recognition accuracy, standard deviation % and optimal dimensions) of LDA, SDA, SMMC, SODA, FME, TR-FSDA and Iterative-SDA with NNC on COIL20 databases. The results in boldface are significantly better than the others.

Methods	LDA	SDA	SMMC	SODA	FME	TR-FSDA	Iterative-SDA
$l=2$	72.93 ± 3.09 (9)	73.99 ± 2.68 (8)	73.76 ± 2.41 (20)	74.98 ± 2.79 (10)	74.87 ± 2.76 (20)	74.75 ± 2.88 (10)	**77.81 ± 3.40 (9)**
$l=3$	77.17 ± 3.15 (18)	79.59 ± 2.81 (7)	79.70 ± 2.79 (8)	81.01 ± 2.50 (9)	80.69 ± 2.35 (20)	80.68 ± 2.38 (8)	**82.92 ± 3.25 (9)**
$l=4$	81.45 ± 3.28 (19)	83.41 ± 2.63 (8)	84.70 ± 3.22 (8)	84.44 ± 2.91 (9)	84.81 ± 2.53 (20)	84.27 ± 2.52 (8)	**86.30 ± 3.13 (9)**
$l=5$	87.03 ± 1.68 (10)	85.80 ± 2.53 (8)	87.42 ± 2.57 (9)	86.74 ± 1.60 (10)	87.35 ± 2.09 (20)	86.42 ± 1.55 (12)	**87.54 ± 2.86 (13)**
$l=6$	89.53 ± 1.74 (11)	87.95 ± 1.36 (12)	89.76 ± 2.06 (8)	89.38 ± 1.54 (10)	88.75 ± 2.12 (20)	89.21 ± 1.53 (12)	**90.11 ± 2.40 (12)**

consuming to collect enough labeled data [12], and automatical labeling methods may not reliable. Futuremore, when the labeled samples are not sufficient, supervised methods can not totally exploit the discriminant information. Semi-supervised learning methods play an important role to address this problem. In recent years, there exist various semi-supervised methods to exploit both of labeled and unlabeled samples.

One strategy for semi-supervised learning is to introduce a geometrical regularizer into a discriminant model by utilizing the local or global structure both of labeled and unlabeled training data. Zhang et al. [13] propose Semi-supervised Dimensionality Reduction (SSDR) to preserve the intrinsic structure of the unlabeled data as well as both the must-link and cannot-link constraints defined on the labeled samples in the projected low-dimensional space. Clearly, SSDR does not consider the locality of data. Cai et al. [14] propose Semi-supervised discriminant analysis (SDA) under the framework of LDA. SDA makes use of labeled data to maximize the separability between different classes and unlabeled data points to explore the local geometric structure of the data. Song et al. [15] propose a semi-supervised dimensionality reduction framework and extend LDA and MMC to Semi-supervised LDA (SSLDA) and Semi-supervised MMC (SMMC), respectively, by maximizing the between-class scatter, and minimizing the within-class scatter and preserving the local structure of a manifold.

Another strategy for semi-supervised learning is to find the optimal labels for all the training samples. Local and global consistency (LGC) [16] and Gaussian fields and harmonic functions (GFHF) [17] propagate the labels from the labeled training data to the unlabeled training data by label fitness and manifold smoothness. Vikas et al. [18] extend regularized least square (RLS) and support vector machine (SVM) into Laplacian RLS (LapRLS) and Laplacian SVM (LapSVM) by adding a geometrical regularizer. In the study [19], Nie et al. argue that the multi-dimensional extension of line RLS can be considered as an out-of-sample extension of LGC/GFHF by enforcing the label prediction matrix to be in the linear subspace spanned by the training data. In order to relax this hard constrain, Nie et al. [19] propose flexible manifold embedding (FME) and Huang et al. [20] propose trace ration based flexible semi-supervised discriminant (TR-FSDA). Though FME and TR-FSDA estimate the label for a new test sample by a projective matrix W and bias b, they can not obtain the low-dimensional representations for original feature. Lu et al. [21] propose cost-sensitive semi-supervised discriminant analysis (CS^3DA) to seek low-dimensional feature representations. CS^3DA first estimates the soft labels for the unlabeled training data and obtains the low dimensional feature by maximizing ration trace of cost-sensitive between-class scatter to within-class scatter. Nie et al. [22] propose semi-supervised orthogonal discriminant analysis (SODA) for dimensionality reduction. SODA also calculates the within-class and between-class scatter matrix by using the predicated labels and obtain the discriminant low dimensional feature with the orthogonal constrain. However, the above mentioned semi-supervised methods all use the original data to construct adjacent graph or to estimate those labels for unlabeled samples. Frequently, the original feature may be contaminated by noise and contains redundant information. Then, the estimated labels or geometric graphs based on the original data often deviate from their true labels or geometric graph, respectively.

A common way to resolve this problem is to take feature selection strategy before subspace learning, as it can select the most informative subset of the original features. Gu et al. [23] propose a unified framework, namely Linear Discriminant Dimensionality Reduction (LDDR), combining feature selection and subspace learning together based on fisher principle. The authors establish the relationship between LDDR model and multi-variate linear regression problem. Relaxing the LDDR model into a $L_{2,1}$ norm constrained least square problem, as it can be solved by accelerated proximal gradient algorithm. In addition, Gu et al. [24] reformulate the subspace learning problem and use $L_{2,1}$ norm on the projection matrix to achieve row sparsity, which leads to selecting relevant features and learning transformation simultaneously. The experimental results [23,24] also illustrate that this strategy is effective. This motives us to estimate the labels F of unlabeled samples in a new feature space, which contains less redundant information and noise. Better established labels of unlabeled can promote to extract better discriminant feature. It motivates us to take an iterative manner to get better results. Similar technology has also been used in the works [25–27]. Specially, Wang et al. [26] propose an iterative fusion approach to graph based semi-supervised learning from multi-views, which joints the process of learning the similarity metric value for graph edge weight and labeling upon the learned metric for semi-supervised learning.

In this paper, our goal is to develop a semi-supervised method Iterative-SDA for feature extraction, which can estimate the labels for unlabeled well. The basic idea of Iterative-SDA is to take an iterative manner to refine the approximation of label information for unlabeled samples. The framework of Iterative-SDA contains the following two steps: (1) Computing the low-dimensional discriminant features; (2) Estimating the label matrix F using low-dimensional features. Those two steps are executed alternatively until the objective function converges. As previous works have demonstrated that integrating the statistically uncorrelated between features is more powerful than the classical feature extraction method in the terms of recognition rata, i.e. uncorrelated linear discrimination analysis (ULDA) [28], local uncorrelated discriminant transform (LUDT) [29] and locally uncorrelated discriminant projections (LUDP) [30]. In the first step, we also impose a statistically uncorrelated constrain to help to remove the redundancy of low-dimensional features. Inspired by Gu et al. [31], we provide an analytical solution by simple eigenvalue decomposition to avoid an iterative process. For the second step, it is an NP hard problem to optimize the binary label matrix F. So, we relax the elements of