Neurocomputing 153 (2015) 96–107

Contents lists available at ScienceDirect

Neurocomputing

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

Image classification via least square semi-supervised discriminant analysis with flexible kernel regression for out-of-sample extension

Mingbo Zhao [a], Bing Li [b,*], Zhou Wu [a], Choujun Zhan [c]

[a] Department of Electronics Engineering, City University of Hong Kong, Hong Kong S.A.R, Kowloon, Hong Kong
[b] School of Economics, Wuhan University of Technology, Wuhan, PR China
[c] Department of Electronic Communication & Software Engineering, Sun Yat-Sen University, Guangzhou, PR China

ARTICLE INFO

Article history:
Received 27 June 2014
Received in revised form
1 September 2014
Accepted 16 November 2014
Communicated by Feiping Nie
Available online 3 December 2014

Keywords:
Dimensionality reduction
Semi-supervised learning
Image classification

ABSTRACT

Semi-supervised dimensionality reduction is an important research topic in many pattern recognition and machine learning applications. Among all the methods for semi-supervised dimensionality reduction, SDA and LapRLS/L are two popular ones. Though the two methods are actually the extensions of different supervised methods, we show in this paper that both SDA and Lap-RLS/L can be unified under a regularized least square framework. In this paper, we propose a new effective semi-supervised dimensionality reduction method for better cope with data sampled from nonlinear manifold. In addition, the proposed method can both handle the regression as well as the subspace learning problem. Theoretical analysis and extensive simulations show the effectiveness of our algorithm. The results in simulations demonstrate that our proposed algorithm can achieve great superiority compared with other existing methods.

1. Introduction

Image classification, such as face recognition, is an important research direction in the field of pattern recognition and machine learning research [1–4]. One of the most successful and well-studied techniques is the appearance-based method. It usually represents an image of $n \times n$ pixels by a $n \times n$-dimensional vector. Usually, most researchers focus on designing a classifier, which can classify image in the $n \times n$-dimensional space directly. However, due to the "curse of dimensionality" [22], dealing with high-dimensional data has always been a major problem for image classification. Finding a low-dimensional representation of high-dimensional space, namely dimensionality reduction, is thus of great practical importance. The goal of dimensionality reduction is to reduce the complexity of input space and embed high-dimensional space into a low-dimensional space while keeping most of the desired intrinsic information [1,2,11,12,29–32]. Among all the dimensionality reduction techniques, Principal Component Analysis (PCA) [3] and linear discriminant analysis (LDA) [4] are two popular methods which have been widely used in many classification applications. PCA pursues the direction of maximum variance for optimal reconstruction. While LDA, as a supervised method, is to find the optimal projection V that maximizes the between-class scatter matrix S_b while minimizes the within-class scatter matrix S_w in the low-dimensional subspace. Due to the utilization of label information, LDA can achieve better classification results than those obtained by PCA if sufficient labeled samples are provided [4].

Though supervised methods generally outperform unsupervised methods, obtaining sufficient number of labeled samples for training can be problematic because labeling large number of samples is time-consuming and costly. On the other hand, unlabeled samples may be abundant and can easily be obtained in the real world. Thus, using semi-supervised learning methods [5–10], which incorporate both labeled and unlabeled samples into learning procedure, has become an effective option instead of only relying on supervised learning. Two well-known semi-supervised learning methods are GFHF [5] and LLGC [6]. These methods work in a tranductive way by propagating the label information from labeled set to unlabeled set via label propagation. But they cannot predict the class labels of new-coming samples hence suffering the out-of-sample problem. In contrast, semi-supervised dimensionality reduction methods can solve this problem. These methods will firstly construct a manifold regularized term to preserve the geometrical structure with both labeled and unlabeled set [13]. Then, they will find a transformed matrix to perform dimensionality reduction by incorporating the manifold regularized term into the original objective function of supervised algorithms. Hence, the new-coming samples can be projected into

* Corresponding author. Tel.: +852 3442 2874.
E-mail addresses: mzhao4@cityu.edu.hk (M. Zhao), lib675@163.com (B. Li), wuzhsky@gmail.com (Z. Wu), zchoujun@gmail.com (C. Zhan).

http://dx.doi.org/10.1016/j.neucom.2014.11.048
0925-2312/

a low-dimensional subspace by using such a transformed matrix and the out-of-sample problem can be solved, which is more practical in real-world applications.

Many semi-supervised dimensionality reduction have been proposed during the past decade. Two widely-used methods are Semi-supervised Discriminant Analysis (SDA) [7] and Linear Laplacian Regularized Least Square (LapRLS/L) [8]. These methods share the same concept for dimensionality reduction, i.e. they first construct the graph Laplacian matrix to approximate the manifold structure by using both labeled and unlabeled samples. They then perform dimensionality reduction by adding the graph Laplacian matrix as a regularized term to the original objective function of LDA and Regularized Least Square (RLS), respectively. Hence both the discriminative structure embedded in the labeled samples and the geometrical structure embedded in both labeled and unlabeled set can be preserved. However, Lap-RLS/L is essentially derived from regression problem instead of classification problem (it can only reduce the dimensionality to c), while SDA is a subspace learning methods which arms to solve classification problem. Though they are initially based on different supervised methods, we show in this paper that both SDA and Lap-RLS/L can be unified under a regularized least square framework. Based on this framework, we then propose a least square version of SDA, referred as LS-SDA, which can both solve regression as well as the subspace learning problem.

In general, most DR methods (e.g. PCA, LDA, Lap-RLS/L, SDA and LS-SDA) are to calculate a linear projection matrix V by assuming the low-dimensional data representation Z is constrained in the linear subspace spanned by the training data matrix X, i.e. $Z=V^TX+b^Te$. Then, the low-dimensional representation Z can be used for faster training and testing in real-world applications. While this linearization constraint provides a simple and effective method to map new samples, Nie et al. argue such constraint is over-strict to fit some data sampled from a nonlinear manifold [9], they then arm to relax this hard constraint in Lap-RLS/L and propose a FME to solve this problem. In fact, both Lap-RLS/L and FME is actually regression problem, i.e. they are to fix a linear model to the labels of labeled samples as well as to preserve the manifold smoothness. Since they are initially derived from regression problem instead of subspace learning problem, they cannot perform subspace learning less than c features. In this paper, to relax this hard constraint in LS-SDA, we add a kernel regression residual term $||V^T\varphi(X)+b^Te-Z||_F^2$ to the reformulated least square framework of SDA. With such relaxation, the proposed method tries to model the mismatch between the low-dimensional data representation Z and the kernel regression function $V^T\varphi(X)+b^Te-Z$ both in the objective function and the constraint, rather than force Z to lie in the subspace spanned by the data matrix X. Hence the proposed method, referred as **L**east **S**quare **S**emi-supervised **D**iscriminant **A**nalysis with **F**lexible Kernel Regression (**LS-SDA/F**), is more flexible and generally suitable to cope with a certain type of nonlinear manifold that is somewhat close to a linear subspace. It is also illustrated that LS-SDA is only a special case LS-SDA/F.

The main contributions of this paper are as follows: (1) We address SDA into a least square framework and proposed a least square version of SDA, which can both solve regression as well as the subspace learning problem; (2) We propose the LS-SDA/L method by relaxing the hard constraint, i.e. $Z=V^TX$, in the least square reformulation of SDA. With this relaxation, LS-SDA/L can better cope with the data sampled from a certain type of nonlinear manifold that is somewhat close to a linear subspace.

This paper is organized as follows: we will give the notations and a brief review of LDA, MR and SDA. In Section 3, we will address SDA under a constrained regularized least square framework. In Section 4, we will present the proposed LS-SDA/F method for semi-supervised dimensionality reduction. Extensive simulations are conducted in Section 5 and the final conclusions are drawn in Section 6.

2. Notations and review of related work

2.1. Linear discriminant analysis (LDA)

The goal of LDA is to seek an optimal projection matrix $V^* \in R^{D\times d}$ that maximizes between-class scatter matrix while minimizes with-class scatter matrix [29]. Let $X_l=\{x_1,x_2,\dots,x_l\}\in R^{D\times l}$ be a set of l samples belonging to c classes, $Y_l=\{y_1,y_2,\dots,y_j\}\in R^{c\times l}$ be the binary label matrix with each column y_j representing the class assignment of x_j, i.e. $y_{ij}=1$, as the class matrix, where $y_{ij}=1$, if x_j belongs to the ith class; $y_{ij}=0$, otherwise, D and c are the numbers of features and classes, respectively. We also denote $G=\{g_1,g_2,\dots,g_l\}=(YY^T)^{-1/2}Y\in R^{c\times l}$ as the scaled class indicator matrix [16,17], where $g_{ij}=1/\sqrt{l_i}$, if x_j belongs to the ith class; $g_{ij}=0$, otherwise. Since YY^T is diagonal matrix, then $GG^T=(YY^T)^{-1/2}YY^T(YY^T)^{-1/2}=I$. Hence assuming the data matrix X_l are centered, the total-class, between-class and within-class scatter matrix S_t, S_b, S_w can be defined as

$$\begin{aligned} S_t &= \textstyle\sum_{i=1}^{c}\sum_{x\in c_i}(x-\mu)(x-\mu)^T = X_lX_l^T \\ S_b &= \textstyle\sum_{i=1}^{c} l_i(\mu_i-\mu)(\mu_i-\mu)^T = X_lG^TGX_l^T \\ S_w &= \textstyle\sum_{i=1}^{c}\sum_{x\in c_i}(x-\mu_i)(x-\mu_i)^T = X_lX_l^T - X_lG^TGX_l^T, \end{aligned} \tag{1}$$

where l_i is the number of samples in the ith class, μ_i is the mean of samples in the ith class, and μ is the mean of all labeled samples. The optimal projection matrix V_{LDA}^* are then formed by eigenvectors corresponding to the d largest eigenvalues of $S_w^{-1}S_b$ or $S_t^{-1}S_b$.

2.2. Manifold regularization (MR)

The MR method [8] extends many existing methods such as least square and SVM to their semi-supervised learning methods by adding a manifold regularized term to preserve the geometrical structure. Take the linear Laplacian regularized Least Square method (referred as Lap-RLS/L) as an example. Let $X=\{X_l,X_u\}=\{x_1,x_2,\dots,x_{l+u}\}\in R^{D\times(l+u)}$ be the data matrix where the first l and the remaining u columns are the labeled and unlabeled samples, respectively. The goal of Lap-RLS/L is to fix a linear model $y_j=V^Tx_j+b^T$ by regressing X on Y and simultaneously to preserve the manifold smoothness embedded in both labeled and unlabeled set, where $V\in R^{D\times d}$ is the projection matrix and $b\in R^{1\times c}$ is the bias term. The objective function of Lap-RLS/L can be given as

$$J(V,b)=\min\sum_{j=1}^{l}\|V^Tx_j+b^T-y_j\|_F^2+\alpha_t\|V\|_F^2+\alpha_m Tr\left(V^TXLX^TV\right), \tag{2}$$

where $L=D-W$ is the graph Laplacian matrix associated with both labeled and unlabeled set [12], W is the weight matrix defined as: $w_{ij}=\exp(-||x_i-x_j||^2/2\sigma^2)$, if x_i is within the k nearest neighbor of x_j or x_j is within the k nearest neighbor of x_i; $w_{ij}=0$, otherwise, D is a diagonal matrix satisfying $D_{ii}=\sum_{j=1}^{l+u}w_{ij}$, α_m and α_t are the two parameters balance the tradeoff between manifold and Tikhonov regularized terms.

2.3. Semi-supervised discriminant analysis (SDA)

Motivated by the manifold regularization (MR), SDA further extends the conventional LDA to preserve the geometric structure by adding a manifold regularized term to the objective function of LDA. The objective function of SDA can be given by:

$$J(V)=\max Tr\left\{\left(V^T\left(S_t+\alpha_tI+\alpha_mXLX^T\right)V\right)^{-1}V^TS_bV\right\}. \tag{3}$$

Then, similar to LDA, the optimal solution of SDA can be obtained by the Generalized Eigenvalue Decomposition as

$$S_bV_{SDA}^*=\left(S_t+\alpha_tI+\alpha_mXLX^T\right)V_{SDA}^*\Lambda. \tag{4}$$