Neurocomputing 74 (2011) 2953–2960

Contents lists available at ScienceDirect

Neurocomputing

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

Kernel machine-based rank-lifting regularized discriminant analysis method for face recognition

Wen-Sheng Chen [a,]*, Pong Chi Yuen [b], Xuehui Xie [a]

[a] College of Mathematics and Computational Science, Shenzhen University, Shenzhen 518060, PR China
[b] Department of Computer Science, Hong Kong Baptist University, Hong Kong, PR China

ARTICLE INFO

Article history:
Received 31 December 2010
Received in revised form
10 March 2011
Accepted 1 April 2011
Communicated by Qi Li
Available online 23 May 2011

Keywords:
Face recognition
Nonlinear problem
Singularity problem
Kernel method
Rank-lifting scheme

ABSTRACT

To address two problems, namely nonlinear problem and singularity problem, of linear discriminant analysis (LDA) approach in face recognition, this paper proposes a novel kernel machine-based rank-lifting regularized discriminant analysis (KRLRDA) method. A rank-lifting theorem is first proven using linear algebraic theory. Combining the rank-lifting strategy with three-to-one regularization technique, the complete regularized methodology is developed on the within-class scatter matrix. The proposed regularized scheme not only adjusts the projection directions but tunes their corresponding weights as well. Moreover, it is shown that the final regularized within-class scatter matrix approaches to the original one as the regularized parameter tends to zero. Two public available databases, namely FERET and CMU PIE face databases, are selected for evaluations. Compared with some existing kernel-based LDA methods, the proposed KRLRDA approach gives superior performance.

1. Introduction

Over the past decade, face recognition has been one of the most active and exciting research topics in computer science and information technology. The challenge to face recognition is that the distributions of face data variations under different poses, illuminations and expressions are very complexity and nonlinear. Thus the linear approach, namely LDA [1], cannot give satisfactory performance. Following the success of applying 'kernel trick' in support vector machine (SVM) [2], many kernel-based discriminant analysis (KDA) methods have been developed and applied to solving nonlinear problems in pattern recognition tasks [3–13]. Kernel methods [15,16] work by implicitly mapping the input samples into a high dimensional feature space, and making the mapped feature space linearly separable. Nevertheless, KDA commonly encounters singularity problem. This problem always occurs when the total number of training samples is smaller than the dimension of feature space. It is known that the dimension of facial pattern vector obtained by vectorizing a facial image is very high. For example, if the resolution of a facial image is 112×92, then the size of facial vector attains 10 304. In such a case, the within-class scatter matrix S_w in mapped feature space becomes singular and direct applying that the KDA approach is impossible. To deal with singularity problem, some kernel-based LDA schemes, such as GDA [5], KDDA [6], K1PRDA [7], etc. methods, have been developed based on different criteria. GDA and KDDA address singularity problem by removing null space from S_w in mapped feature space. However, the null space of S_w contains much useful discriminant information for pattern classification [19]. Furthermore, it is not reasonable that the kernel matrix in GDA method is assumed to be nonsingular. K1PRDA is a regularization method, which uses three-to-one regularization technique [20] to guarantee full rank of within-class scatter matrix S_w. K1PRDA finds the optimal projection only by tuning the weights (eigenvalues) of directions (eigenvectors) while all directions are kept with no variations. Recently, based on rank-lifting technique, a two-step regularization Fisher discriminant analysis (2SRFD) [21] approach has been presented to solve singularity problem. But it just a linear method and its second regularization step must be performed after the first step.

This paper, motivated by our previous 2SRFD [21] method, proposes a novel kernel machine-based rank-lifting regularized discriminant analysis (KRLRDA) method to tackle nonlinear problem and singularity problem in face recognition. Our regularization technique also includes two regularization stages. A new rank-lifting theorem using linear algebraic theory is first shown. Based on this theorem, a rank-lifting parameter t is exploited to increase the rank of matrix Φ_w ($S_w^{\varphi}=\Phi_w\Phi_w^T$) and then a $d\times N$ column full rank matrix $\tilde{\Phi}_w$ is obtained, where d is the dimension

* Corresponding author. Tel.: +86 755 26538946; fax: +86 755 26538959.
E-mail addresses: chenws@szu.edu.cn (W.-S. Chen), pcyuen@comp.hkbu.edu.hk (P.C. Yuen).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.04.019

of original feature space, N is the total number of training data. The rank-lifted within-class scatter matrix $\tilde{S}_w^{\varphi t}$ is subsequently derived. In most cases, the matrix $\tilde{S}_w^{\varphi t}$ is full rank and our KRLRDA approach can be performed directly. If $\tilde{S}_w^{\varphi t}$ is still not invertible, the second regularization step shall employ the regularization technique proposed in the literature [20] with regularized parameter s. After above two regularization steps, the complete regularized matrix $\tilde{S}_w^{\varphi ts}$ is obtained. It is demonstrated that the final regularized within-class scatter matrix $\tilde{S}_w^{\varphi ts}$ approaches to original within-class scatter matrix $\tilde{S}_w^{\varphi}$ as the regularization parameters t and s tend to zero simultaneously. A novel KRLRDA method is subsequently developed and applied to face recognition.

In our KRLRDA approach, the first regularization stage adjusts not only the directions of projection but the magnitudes (weight) in each direction as well. The eigenvalues can be considered as the 'weight' in the corresponding direction (eigenvector). In the second regularization stage, only the weights of directions are adjusted while the projection directions are kept invariant [20]. Especially, in most cases, the first regularization stage using our rank-lifting strategy is sufficient to guarantee the full rank of within-class scatter matrix, while the second regularization step is just standby in case the first stage fails to work.

Two public available databases, namely FERET and CMU PIE face databases, are selected for evaluations. Compared with some existing kernel-based LDA methods, experimental results show that the proposed KRLRDA approach gives superior performance.

The rest of this paper is organized as follows. Detail theoretical analysis and KRLRDA algorithm design are proposed in Section 2. The experimental results and computational complexity are reported in Section 3. Finally, Section 4 draws the conclusions.

2. Proposed KRLRDA method

In this section, a new kernel machine-based rank-lifting RDA method is proposed. Details are discussed below.

2.1. Some notations

Let d be the dimension of original feature space and C be the number of sample classes. The total original sample set $X=\bigcup_{j=1}^{C}X_j$, where the jth class $X_j=\{x_i^j\}_{i=1}^{N_j}$ contains N_j training samples. Assume $N(=\sum_{j=1}^{C}N_j)$ is the number of total training samples and $\varphi(x): x\in R^d\rightarrow\varphi(x)\in\mathcal{F}$ is the kernel nonlinear mapping, where $\mathcal{F}$ is the mapped feature space with dimension $df(=dim\,\mathcal{F})$. The total mapped training sample set and the jth mapped class are given by $\varphi(X)=\bigcup_{j=1}^{C}X_j$ and $\varphi(X_j)=\{\varphi(x_i^j)\}_{i=1}^{N_j}$, respectively. The mean of the mapped sample class $\varphi(X_j)$ and the global mean of the total mapped samples $\varphi(x)$ are given by $m_j=(1/N_j)\sum_{i=1}^{N_j}\varphi(x_i^j)$ and $m=(1/N)\sum_{j=1}^{C}\sum_{i=1}^{N_j}\varphi(x_i^j)$, respectively. In feature space $\mathcal{F}$, two scatter matrices, namely within-class and between-class matrices, are defined, respectively, as follows:

$$S_w^{\varphi}=\frac{1}{N}\sum_{j=1}^{C}\sum_{i=1}^{N_j}(\varphi(x_i^j)-m_j)(\varphi(x_i^j)-m_j)^T=\Phi_w\Phi_w^T,$$

$$S_b^{\varphi}=\frac{1}{N}\sum_{j=1}^{C}N_j(m_j-m)(m_j-m)^T=\Phi_b\Phi_b^T,$$

where $\Phi_w=\sqrt{(1/N)}[\varphi(x_i^j)-m_j]_{i=1,\ldots,N_j}^{j=1,\ldots,C}$ is a $df\times N$ matrix, and $\Phi_b=[\sqrt{(N_j/N)}(m_j-m)]_{j=1,\ldots,C}$ is a $df\times C$ matrix. In addition, we define a $df\times N$ mapped training sample matrix $\Phi\times\Phi=[\varphi(x_i^j)]_{i=1,\ldots,N_j}^{j=1,\ldots,C}\in R^{df\times N}$.

The Fisher index $J_\Phi(w)$ in mapped feature space $\mathcal{F}$ is defined by

$$J_\varphi(w)=\frac{w^TS_b^{\varphi}w}{w^TS_w^{\varphi}w}, \tag{1}$$

where $w\in\mathcal{F}$.

According to the Mercer kernel function theory [14], any solutions $w\in\mathcal{F}$ belong to the span of all training patterns in $\mathcal{F}$. Hence there exists a coefficient column vector $\tilde{w}\in R^N$ such that $w=\Phi\tilde{w}$. Substituting $w=\Phi\tilde{w}$ into (1), the Fisher criterion function in the mapped feature space $\mathcal{F}$ can be rewritten as follows:

$$J_\varphi(\tilde{w})=\frac{\tilde{w}^T\tilde{S}_b^{\varphi}\tilde{w}}{\tilde{w}^T\tilde{S}_w^{\varphi}\tilde{w},} \tag{2}$$

where

$$\tilde{S}_b^{\varphi}=\Phi^T\Phi_b\Phi_b^T\Phi \quad\text{and}\quad \tilde{S}_w^{\varphi}=\Phi^T\Phi_w\Phi_w^T\Phi.$$

KDA aims to find an optimal projection $\tilde{w}_{opt}$ which maximizes the Fisher criterion function (2), namely,

$$\tilde{w}_{opt}=\arg\max_{\tilde{w}}J_\varphi(\tilde{w}).$$

Above problem is equivalent to solving the following eigensystem:

$$(\tilde{S}_w^{\varphi})^{-1}\tilde{S}_b^{\varphi}\tilde{w}=\lambda\tilde{w}. \tag{3}$$

However, the matrix $\tilde{S}_w^{\varphi}$ is not invertible when singularity problem occurs. In this case, the kernel-based LDA (KDA) method cannot be used directly. In turn, we propose to regularize the $\tilde{S}_w^{\varphi}$.

2.2. Theoretical analysis

To overcome singularity problem of KDA, we shall use regularization technique. The proposed regularization strategy involves two regularization stages, namely Rank-lifting stage and Three-to-one regularization stage. Below is the detail theoretical analysis on our KRLRDA approach.

In the first step, we shall need the following theorem.

Theorem 1. *Given an original vector set in* R^d *as* $\bigcup_{j=1}^{C}\{\alpha_i^j|i=1,2,\ldots,N_j\}$, *let* $M_j=(1/N_i)\sum_{i=1}^{N_j}\alpha_i^j$, *then for any constant* t $(t\neq 0)$, *the following modified vector set:*

$$\bigcup_{j=1}^{C}\{\alpha_i^j-M_j+tM_j|i=1,2,\ldots,N_j\}$$

is equivalent to the original one.

Proof. For $j=1,2,\ldots,C$, let $A_j=\bigcup_{i=1}^{N_j}\{\alpha_i^j\}$ and $B_j=\bigcup_{i=1}^{N_1}\{\alpha_1^j-M_j+tM_j\}$. It is easy to see that set B_j can be expressed by the linear combination of set A_j. In the following, it will be shown that A_j can also be expressed by the linear combination of B_j. To this end, assume $\beta_i^j=\alpha_i^j-M_j+tM_j$, $i=1,2,\ldots,N_j$, and then $B_j=\bigcup_{i=1}^{N_j}\{\beta_i^j\}$. Since

$$\sum_{i=1}^{Nj}\beta_i^j=\sum_{j=1}^{Nj}(\alpha_i^j-M_j+tM_j)=N_jt\cdot M_j,$$

and $t\neq 0$, it yields that $M_j=(\sum_{i=1}^{N_j}\beta_i^j)/(N_jt)$, and then $\alpha_i^j=\beta_i^j+(1-t)(\sum_{k=1}^{N_j}\beta_i^j)/(N_jt), i=1,2,\ldots,N_j$. Therefore, set A_j is the linear combination of set B_j. Above analysis demonstrates that A_j is equivalent to B_j. It means that the set $\bigcup_{j=1}^{C}\{\alpha_i^j|i=1,2,\ldots,N_j\}$ is