Neurocomputing 000 (2018) 1–10

Contents lists available at ScienceDirect

Neurocomputing

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

Multiple kernel locality-constrained collaborative representation-based discriminant projection for face recognition

Zhichao Zheng, Huaijiang Sun*, Guoqing Zhang

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

ARTICLE INFO

Article history:
Received 14 January 2017
Revised 22 August 2018
Accepted 27 August 2018
Available online xxx

Communicated by Prof. Zidong Wang

Keywords:
Collaborative representation
Dimensionality reduction
Multiple kernel learning
Trace ratio optimization
Locality constraint

ABSTRACT

Collaborative representation-based classifier (CRC) has achieved superior classification performance in the field of face recognition. However, the performance of CRC will degrade significantly when facing nonlinear structural data. To address this problem, many kernel CRC (KCRC) methods have been proposed. These methods usually use a predetermined kernel function which is difficult to be selected. In addition, how to select appropriate parameters remains a challenging problem. Hence, multiple kernel technology (MKL) is applied on CRC, which called MK-CRC. However, it only considers the representation errors while ignoring the class label information in the training process. In this paper, we propose a multiple kernel locality-constrained collaborative representation-based classifier (MKLCRC) which is the multiple kernel extension of CRC and considers the local structures of data. Based on the classification rule of MKLCRC, we propose a dimensionality reduction (DR) method called multiple kernel locality-constrained collaborative representation-based discriminant projection (MKLCR-DP). The goal of MKLCR-DP is to learn a projection matrix and a set of kernel weights to generate a low-dimensional subspace where the between-class reconstruction errors are maximized and the within-class reconstruction errors are minimized. Thus MKLCRC can achieve better performance in this low-dimensional subspace. The proposed method can be efficiently optimized with the trace ratio optimization. Experiments on AR, extended Yale B, FERET, CMU PIE and LFW face databases demonstrate that our method outperforms related state-of-the-art algorithms.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Face recognition is an important and interesting topic in pattern recognition areas during the past 20 years, and has been widely used in many fields such as commercial and security [1,2]. It is well known that the dimension of face image is usually very high. For example, a 200-by-200 pixel face image can be stretched as a 40,000 dimensional vector. The high dimensional data usually contains redundant information and makes the storage space high. On the other hand, the high dimension increases the computational cost. A common way to solve these problems is using dimensionality reduction (DR) methods which can find a compact representation and enhance the discrimination of face image in the low-dimensional subspace. One of the popular DR methods is principle component analysis (PCA) [3], which maximizes the variance of projected data. Another is linear discriminant analysis (LDA) [4]. It attempts to project the data into a low-dimensional subspace where the same-labeled data is shrunk and the different-labeled data is distinguished. Both PCA and LDA are linear DR methods that assume the distributions of data are linear. However, face recognition is a nonlinear problem with variations in illumination, facial expression, occlusion and view angle [5]. Hence, the kernel trick [6] is used to solve this problem which uses kernel function for implicitly mapping data from original space into a high-dimensional space, called Reproducing Kernel Hilbert space, where the nonlinear problems may be linearly separable. The popular kernel functions include Gaussian kernel function, linear kernel function, and sigmoid kernel function. Kernel PCA (KPCA) [7] and Kernel LDA (KLDA) [8] are generated by applying the kernel method to PCA and LDA, respectively.

Recently, some studies have found that the sparse representation-based classification (SRC) is an effective tool to deal with the face recognition problem [9,10]. In SRC, the test sample can be presented linearly by the whole over training set under the sparsity constraint, then classification can be done by choosing the minimum reconstruction error. Solving the ℓ_1-norm minimization problem makes SRC very expensive [11]. Zhang et al. [12] argued that the collaborative representation principle, i.e., using the whole training set to represent the test

* Corresponding author.
E-mail addresses: zheng_zhichaoX@163.com (Z. Zheng), sunhuaijiang@njust.edu.cn (H. Sun), xiayang14551@163.com (G. Zhang).

https://doi.org/10.1016/j.neucom.2018.08.032
0925-2312/© 2018 Elsevier B.V. All rights reserved.

Please cite this article as: Z. Zheng et al., Multiple kernel locality-constrained collaborative representation-based discriminant projection for face recognition, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.08.032

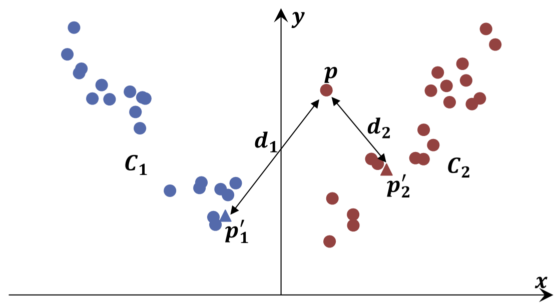

Fig. 1. The case of two-dimension. There is a test sample $\boldsymbol{p}$ whose label is C_2. $\boldsymbol{p}_1'$ and $\boldsymbol{p}_2'$ are the projections of $\boldsymbol{p}$ in C_1 class and C_2 class. The Euclidean distances between $\boldsymbol{p}$ with $\boldsymbol{p}_1'$ and $\boldsymbol{p}_2'$ are d_1 and d_2. In the classification rule of CRC, the classification process of $\boldsymbol{p}$ can be viewed as choosing the minimum value of d_1 and d_2. Therefore, we can project these samples to a low-dimensional subspace where d_2 is smaller and d_1 is larger. These two mechanisms guarantee that $\boldsymbol{p}$ is more tend to be correctly classified.

sample, plays a more important role than sparsity. Furthermore, the sparsity constraint is only helpful to robustness, but not accuracy. Then, a collaborative representation-based classification with regularized least square (CRC) was proposed in [12]. CRC replaces sparsity constraint with least square constraint, and has significantly less complexity than SRC but competitive in classification performance. Further, Liu et al. [13] proposed the kernel collaborative representation-based classification (KCRC). Wei et al. [14] claimed that DR methods should be designed according to the classifiers, thus, they presented a DR method called kernel locality-constrained collaborative representation based discriminant analysis (KLCR-DA).

In recent years, several researches have reported that Multiple Kernel Learning (MKL) [15] which using multiple different kernels instead of single kernel improves the classification performance [16,17]. Lin et al. [18] incorporated MKL into DR and proposed multiple kernel learning for dimensionality reduction (MKL-DR). To speed up MKL-DR, Liu et al. [19] proposed a multiple kernel dimensionality reduction algorithm based on spectral regression (MKL-SR) which transforms the eigenvalue decomposition problem into a linear regression problem. Liu et al. [20] proposed a multiple kernel collaborative representation based classification (MK-CRC) method which only considers global reconstruction error to optimize the kernel weights, while ignoring the class label information.

Inspired by the above concerns, we propose a novel DR method called multiple kernel locality-constrained collaborative representation based discriminant projection (MKLCR-DP) which is designed according to the decision rule of our proposed multiple kernel locality-constrained collaborative representation-based classifier (MKLCRC). The goal of MKLCR-DP is to learn a projection matrix and a set of kernel weights such that the between-class reconstruction errors are maximized and the within-class reconstruction errors are minimized in the low-dimensional subspace. MKLCRC can achieve a good performance in this subspace. The reason is shown in Fig 1. In addition, we utilize the local structures of data which contain more information than the global structure of data. Finally, the projection matrix and the kernel weights are optimized by a trace ratio optimization algorithm where the trace ratio maximization problem is converted to a single trace maximization problem. The trace ratio optimization algorithm is more effective and efficient than generalized eigenvalue decomposition method. The process of MKLCR-DP is shown in Fig 2.

The remainder of this paper is organized as follows. Section 2 presents some related work. Details of MKLCRC and MKLCR-DP are given in Section 3 and Section 4. Experimental results are discussed in Section 5, and Section 6 concludes the paper.

The key acronyms of our paper are shown in Table 1.

2. Related work

2.1. Collaborative representation-based classification

Suppose that there is a training set $\boldsymbol{X}$ containing total $\boldsymbol{C}$ classes and the ith class has n_i samples. $\boldsymbol{X}$ can be formulated as $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_C\} \in \mathbb{R}^{m\times n}$ where m is the dimension of each training sample and $n = \sum_{i=1}^{k} n_i$. $\boldsymbol{X}_i = \{\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{in_i}\} \in \mathbb{R}^{m\times n_i}$ denotes the training set associated with the ith class and x_{ij} stands for the jth training sample in class i. In CRC, the test image $\boldsymbol{y}$ should be presented over the whole training set $\boldsymbol{X}$ but not subset $\boldsymbol{X}_i$. Therefore, $\boldsymbol{y}$ can be presented as $\boldsymbol{y} \approx w_1\boldsymbol{x}_1 + w_2\boldsymbol{x}_2 + \cdots + w_n\boldsymbol{x}_n = \boldsymbol{Xw}$, where $\boldsymbol{w} = [w_1, w_2, ..., w_n]^T$ is the collaborative representation coefficients. The vector $\boldsymbol{w}$ can be found by solving the constrained optimization problem:

$$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \|\boldsymbol{w}\|_2^2 \ s.t. \|\boldsymbol{y} - \boldsymbol{Xw}\|_2^2 \le \epsilon, \tag{1}$$

where ϵ is a small error constant. After Lagrangian formulation, a general model of CRC is formulated as

$$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \{\|\boldsymbol{y} - \boldsymbol{Xw}\|_2^2 + \lambda\|\boldsymbol{w}\|_2^2\}, \tag{2}$$

in which λ is the regularization parameter. $\widehat{\boldsymbol{w}}$ can be analytically derived as

$$\widehat{\boldsymbol{w}} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}. \tag{3}$$

where $\boldsymbol{I}$ is a identity matrix. The residuals of different classes can be computed as

$$R_i(\boldsymbol{y}) = \|\boldsymbol{y} - \boldsymbol{X}\delta_i(\widehat{\boldsymbol{w}})\|_2^2 / \|\delta_i(\widehat{\boldsymbol{w}})\|_2^2, \ i = 1, 2, ..., C, \tag{4}$$

where $\delta_i(\cdot)$ is a characteristic function which selects the coefficients associated with the ith class. Then $\boldsymbol{y}$ is assigned to the class with minimal error.

2.2. Kernel collaborative representation-based classification

The performance of CRC is restricted facing the non-linear structure data. To address this problem, the data is non-linearly transformed to the high-dimensional space where CRC is applied on the new features. Using the non-linear mapping ϕ, $\boldsymbol{X}$ can be expressed as $\boldsymbol{\Phi} = \{\phi(\boldsymbol{x}_1), \phi(\boldsymbol{x}_2), ..., \phi(\boldsymbol{x}_n)\} \in \mathbb{R}^{d\times n}$ where d is the dimension of high-dimensional space and $d \gg m$. The model of CRC in high-dimensional space can be formulated as

$$\arg\min_{\boldsymbol{w}} \|\boldsymbol{w}\|_2^2 \ s.t. \|\phi(\boldsymbol{y}) - \boldsymbol{\Phi w}\|_2^2 \le \epsilon, \tag{5}$$

However, the dimension of high-dimensional space is very high or possibly infinite which makes optimization is complicated in CRC. Hence, reducing the dimensionality of high-dimensional space is necessary. The projection matrix $\boldsymbol{P}$ can be defined and constructed by utilizing the methodology in KPCA [7] and KLDA [8]. Projection on high-dimensional space can be written as

$$\boldsymbol{P}^T\phi(\boldsymbol{y}) = \boldsymbol{P}^T\boldsymbol{\Phi w}. \tag{6}$$

In fact, each column of $\boldsymbol{P}$ is the linear combination of samples in high-dimensional space. Namely, $\boldsymbol{P}_i = \sum_{j=1}^{n} t_{i,j}\phi(\boldsymbol{x}_j) = \boldsymbol{\Phi t}_i$. $\boldsymbol{t}_i = [t_{i,1}, t_{i,2}, ..., t_{i,n}]^T$ is the transformation vector corresponding to the ith projection vector. Therefore, $\boldsymbol{P}$ can be expressed as

$$\boldsymbol{P} = \boldsymbol{\Phi T}. \tag{7}$$

$\boldsymbol{T} = [\boldsymbol{t}_1, \boldsymbol{t}_2, ..., \boldsymbol{t}_L] \in \mathbb{R}^{n\times L}$ is the transformation matrix. Then we put Eq. (7) into Eq. (6):

$$\boldsymbol{T}^T\boldsymbol{\Phi}^T\phi(\boldsymbol{y}) = \boldsymbol{T}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi w}, \tag{8}$$

Download English Version:

https://daneshyari.com/en/article/11021142

Download Persian Version:

https://daneshyari.com/article/11021142

Daneshyari.com