Pattern Recognition Letters 73 (2016) 83–90

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Optimized projection for Collaborative Representation based Classification and its applications to face recognition☆

Jun Yin*, Lai Wei, Miao Song, Weiming Zeng

College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

ARTICLE INFO

Article history:
Received 4 September 2015
Available online 2 February 2016

Keywords:
Dimensionality reduction
Collaborative representation
Face recognition
Discriminant analysis

ABSTRACT

Collaborative Representation based Classification (CRC) is powerful for face recognition and has lower computational complexity than Sparse Representation based Classification (SRC). To improve the performance of CRC, this paper proposes a new dimensionality reduction method called Optimized Projection for Collaborative Representation based Classification (OP-CRC), which has the direct connection to CRC. CRC uses the minimum reconstruction residual based on collaborative representation as the decision rule. OP-CRC is designed according to this rule. The criterion of OP-CRC is maximizing the collaborative representation based between-class scatter and minimizing the collaborative representation based within-class scatter in the transformed space simultaneously. This criterion is solved by iterative algorithm and the algorithm converges fast. CRC performs very well in the transformed space of OP-CRC. Experimental results on Yale, AR, FERET, CMU_PIE and LFW databases show the effectiveness of OP-CRC in face recognition.

1. Introduction

Face recognition has been a research focus of pattern recognition and computer vision in the past few decades [1]. The dimensionality of face image is often very high. This increases the computational cost and possibly causes “curse of dimensionality”. To solve this problem, numerous dimensionality reduction methods were proposed. These methods reduce the dimension of face image and achieve the discriminating features.

Principal Component Analysis (PCA) [2] and Linear Discriminant Analysis (LDA) [3] are two classical dimensionality reduction methods. PCA finds a mapping by maximizing the variance, and LDA derives a projection from maximizing the between-class scatter and minimizing the within-class scatter simultaneously. However, these two methods fail to discover the manifold structure of data. Manifold learning methods such as Local Preserving Projection (LPP) [4], Neighborhood Preserving Embedding (NPE) [5] and Isometric Projection (IsoProjection) [6] overcome this limitation. They assume the data lie in a low dimensional manifold of the high dimensional space and discover this manifold. LPP, NPE and IsoProjection are unsupervised methods. To further increase the discriminating ability, some supervised manifold learning methods were developed. Local Discriminant Embedding (LDE) [7], Marginal Fisher Analysis (MFA) [8], Locality Preserving Discriminant Projections (LPDP) [9] and Discriminant Simplex Analysis (DSA) [10] are the representative methods. Moreover, a Graph Embedding framework [8] was proposed, the manifold learning methods can be reformulated in this framework.

Recently, a Sparse Representation Based Classification (SRC) [11] method was proposed for face recognition. In SRC, a test sample is represented as a sparse linear combination of all the training samples and this test sample is assigned to the class with the smallest reconstruction residual. Kernel Sparse Representation based Classification (KSRC) [12, 13] performs SRC in a higher dimensional space by using kernel function. It is more effective for nonlinear data. In SRC and KSRC, sparse representation is used for classification. Sparse representation is also used for dimensionality reduction in some methods such as Sparsity Preserving Projection (SPP) [14] and Sparse Neighborhood Preserving Embedding (SNPE) [15]. They construct L1-gragh using sparse representation coefficients, and then obtain the low dimensional features by Graph Embedding. Discriminant Sparsity Preserving Embedding (DSPE) [16] and Discriminant Sparse Neighborhood Preserving Embedding (DSNPE) [17] construct the discriminant L1-graph by introducing class information into sparse representation. Furthermore, Yang et al. proposed SRC steered Discriminative Projection (SRC-DP) [18] and Lu and Huang proposed Optimized Projection for Sparse Representation based Classification (OP-SRC) [19]. SRC-DP and OP-SRC both try to find the low dimensional features which are optimal for SRC. SRC could perform better in the transformed space of these two methods.

☆ This paper has been recommended for acceptance by J. Yang.

* Corresponding author. Tel.: +86 2138282823.
E-mail address: yinjun8429@163.com, junyin@shmtu.edu.cn (J. Yin).

http://dx.doi.org/10.1016/j.patrec.2016.01.012

Very recently, Zhang et al. indicated that it is collaborative representation not L1-norm sparsity which makes SRC powerful and developed a face recognition method called Collaborative Representation based Classification (CRC) [20]. For a test sample, CRC also uses all the training samples to represent it. The main difference of SRC and CRC is the regularized term. L2-norm but not L1-norm is adopted in the objective function of CRC, this makes CRC have significantly lower complexity than SRC. Kernel Collaborative Representation Classification (KCRC) [21] maps the data into a higher dimensional space where different classes are more separable and then performs CRC in this new space. Using collaborative representation, Yang et al. proposed a dimensionality reduction method called Collaborative Representation based Projections (CRP) [22] for face recognition. CRP constructs L2-graph using collaborative representation coefficients and reduces dimensionality by graph embedding. It could preserve the collaborative representation based reconstruction relationship and is faster than dimensionality reduction methods based on sparse representation.

Although CRP utilizes collaborative representation, it has no direct connection to CRC. To improve the classification performance of CRC, we should strengthen the connection of the dimensionality reduction method and CRC. In this paper, a new dimensionality reduction method named Optimized Projection for Collaborative Representation based Classification (OP-CRC) is developed. OP-CRC is designed according to the classification criterion of CRC. It seeks a projection that maximizes the collaborative representation based between-class reconstruction residual and minimizes the collaborative representation based within-class reconstruction residual simultaneously. In the transformed space of OP-CRC, CRC could achieve the optimum classification performance.

The remaining of the paper is organized as follows: Section 2 briefly reviews CRC. Section 3 gives the proposed OP-CRC method. Section 4 presents the experimental results to demonstrate the effectiveness of OP-CRC. Finally, the conclusions are provided in Section 5.

2. Collaborative Representation based Classification

Suppose there are n training samples from c classes and the ith class has n_i training samples. Let $X = [X_1, X_2, \ldots, X_c] \in \mathbb{R}^{m\times n}$ be the training sample matrix, where $X_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,n_i}] \in \mathbb{R}^{m\times n_i}$ is the training sample matrix of the ith class and m is the dimension of the sample.

In CRC, a test sample y is collaboratively represented by all the training samples. The collaborative representation coefficient vector $\hat{\alpha}$ is achieved by solving the following regularized minimization problem:

$$\hat{\alpha} = \arg\min_{\alpha}\{\|y - X\alpha\|_2^2 + \eta\|\alpha\|_2^2\} \tag{1}$$

where η is the regularization parameter. Unlike SRC, CRC uses L2-norm instead of L1-norm in the regularized term. The solution $\hat{\alpha}$ of Eq. (1) can be easily derived as

$$\hat{\alpha} = (X^TX + \eta\cdot I)^{-1}X^Ty \tag{2}$$

With $\hat{\alpha}$, the collaborative representation based reconstruction residual of each class can be calculated. In SRC, the test sample y is assigned to the class which has the minimum residual. Considering that L2-norm of $\hat{\alpha}$ contains some discrimination information, CRC uses the following classification criterion:

$$identity(y) = \arg\min_{i}\{\|y - X_i\hat{\alpha}_i\|_2/\|\hat{\alpha}_i\|_2\} \tag{3}$$

where $\hat{\alpha}_i$ is the coefficient vector associated with the ith class.

3. Optimized projection for Collaborative Representation based Classification

CRC performs very well in face recognition and it is computationally more efficient than SRC. To enhance the performance of CRC, we propose Optimized Projection for Collaborative Representation based Classification (OP-CRC). The optimum features for CRC could be achieved by the transformation of OP-CRC. Therefore, CRC can perform better in the transformed space of OP-CRC.

3.1. The OP-CRC algorithm

Let $P \in \mathbb{R}^{m\times d}$ be the optimized projection matrix. Each training sample $x_{ij} \in \mathbb{R}^m$ is mapped into $y_{ij} = P^Tx_{ij} \in \mathbb{R}^d$ and the training sample matrix $X \in \mathbb{R}^{m\times n}$ is mapped into $Y = P^TX \in \mathbb{R}^{d\times n}$. In the transformed space, for a training sample y_{ij} from the ith class, use the remaining training samples to linearly represent it. The collaborative representation coefficient vector $\hat{\alpha}$ is obtained by solving the regularized minimization problem in Eq. (1). Let $\hat{\alpha}_k$ be the coefficient vector associated with the kth class. The collaborative representation based reconstruction residual of the kth class is

$$R_k(y_{ij}) = \left\|y_{ij} - Y\hat{\alpha}_k\right\|_2 \tag{4}$$

To make CRC perform well in the transformed space, for the sample y_{ij}, the residual of the ith class should be as small as possible and the residuals of other classes should be as big as possible. Using the residual of each class, we define the collaborative representation based within-class scatter and the collaborative representation based between-class scatter. The collaborative representation based within-class scatter is defined as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(R_i(y_{ij}))^2\\
&= \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\left\|y_{ij} - Y\hat{\alpha}_i\right\|_2^2\\
&= \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(y_{ij} - Y\hat{\alpha}_i)^T(y_{ij} - Y\hat{\alpha}_i)\\
&= tr\left(\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}P^T(x_{ij} - X\hat{\alpha}_i)(x_{ij} - X\hat{\alpha}_i)^TP\right)\\
&= tr\left(P^TS_wP\right),
\end{aligned} \tag{5}$$

where

$$S_w = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(x_{ij} - X\hat{\alpha}_i)(x_{ij} - X\hat{\alpha}_i)^T \tag{6}$$

S_w is called the collaborative representation based within-class scatter matrix. The collaborative representation based between-class scatter is defined as

$$\begin{aligned}
&\frac{1}{n(c-1)}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\sum_{k\neq i}(R_k(y_{ij}))^2\\
&= \frac{1}{n(c-1)}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\sum_{k\neq i}\left\|y_{ij} - Y\hat{\alpha}_k\right\|_2^2\\
&= \frac{1}{n(c-1)}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\sum_{k\neq i}(y_{ij} - Y\hat{\alpha}_k)^T(y_{ij} - Y\hat{\alpha}_k)\\
&= tr\left(\frac{1}{n(c-1)}\sum_{i=1}^{c}\sum_{j=1}^{n_i}\sum_{k\neq i}P^T(x_{ij} - X\hat{\alpha}_k)(x_{ij} - X\hat{\alpha}_k)^TP\right)\\
&= tr\left(P^TS_bP\right),
\end{aligned} \tag{7}$$