Knowledge-Based Systems 70 (2014) 212–220

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Kernel locality-constrained collaborative representation based discriminant analysis

Lai Wei [a,*], Feifei Xu [b], Jun Yin [a], Aihua Wu [a]

[a] *Department of Computer Science, Shanghai Maritime University, Haigang Avenue 1550, Shanghai, PR China*
[b] *Department of Computer Science, Shanghai University of Electric Power, Pingliang Road 2103, Shanghai, PR China*

ARTICLE INFO

Article history:
Received 27 October 2013
Received in revised form 29 June 2014
Accepted 29 June 2014
Available online 8 July 2014

Keywords:
Feature extraction
Sparse representation
Collaborative representation
Local structures
Kernel methods

ABSTRACT

Collaborative representation based classifier (CRC) has been successfully applied to pattern classification. However, CRC may not be able to identify the data with highly nonlinear distribution as a linear algorithm. In this paper, we first propose a kernel locality-constrained collaborative representation based classifier (KLCRC). KLCRC is a nonlinear extension of CRC, and it introduces the local structures of data sets into collaborative representation methods. Since the kernel feature space has a very high (or possibly infinite) dimensionality, we present a dimensionality reduction method (termed kernel locality-constrained collaborative representation based discriminant analysis, KLCR-DA) which can fit KLCRC well. KLCR-DA seeks a subspace in which the between-class reconstruction residual of a given data set is maximized and the within-class reconstruction residual is minimized. Hence, KLCRC can achieve better performances in the projected space. Extensive experimental results on AR, the extended Yale B, FERET face image databases and HK PloyU palmprint database show the superiority of KLCR-DA in comparison to the related methods.

1. Introduction

In computer vision fields, many applications, such as appearance-based image recognition, often confront high-dimensional data samples [1,2]. The high-dimensional data usually leads to the inefficiency of many practical data processing techniques and may even degrade the performances of many classifiers. Hence, it is desired to consider methods of feature extraction (or dimensionality reduction) which are able to find the low-dimensional and compact representations for the high-dimensional data samples.

In the past few decades, subspace learning [3,4] based feature extraction algorithms have become some of the most popular ones. The two most widely-used subspace learning methods are principal component analysis (PCA) [5] and linear discriminant analysis (LDA) [6]. PCA and LDA are both linear dimensionality reduction algorithms. They assume that the distributions of data sets are globally linear. However, high-dimensional data sets usually have nonlinear structures in practical. Therefore, kernel tricks [7] are used to generalize PCA and LDA to be nonlinear algorithms. Kernel PCA [8] and Kernel LDA [9] have been proved to be much more efficient in nonlinear feature extraction tasks than PCA and LDA. On the other hand, a family of manifold learning related subspace learning methods which explore and use the local structures of data sets have been proposed. For example, He et al. proposed two well-known algorithms, locality preserving projections (LPP) [10] and neighborhood preserving projection (NPE) [11] which can be regarded as the linear versions of Laplacian Eigenmap (LE) [12] and locally linear embedding (LLE) [13] respectively. Yan et al. presented a marginal Fisher analysis (MFA) [14] algorithm and formulated many existing subspace learning algorithms into a graph-embedding framework. Sugiyama proposed a local Fisher discriminant analysis [15] by introducing the local structure information of data sets into LDA.

In recent years, the idea of sparse representation is used to design some feature extraction methods. Wright et al. presented a sparse representation-based classification (SRC) method [16] for face recognition. In SRC, a query image is sparsely reconstructed by a group of training images. The obtained reconstruction coefficients can be regarded as the affinities between the query image and different training images. Based on SRC, Qiao et al. constructed a kind of l_1-norm affinity graph and presented a sparsity preserving projection (SPP) method [17]. SPP seeks a subspace in which the sparse reconstructive relationship of original data points is preserved. Gui et al. introduced the class information of data sets into

* Corresponding author. Tel.: +86 21 38282826; fax: +86 21 38282800.
E-mail addresses: weilai@shmtu.edu.cn (L. Wei), xufeifei1983@hotmail.com (F. Xu), junyin@shmtu.edu.cn (J. Yin), ahwu@shmtu.edu.cn (A. Wu).

http://dx.doi.org/10.1016/j.knosys.2014.06.027
0950-7051/

SPP and presented a discriminant sparsity neighborhood preserving embedding (DSNPE) [18]. Yang et al. claimed that dimensionality reduction algorithms should be designed according to the classifiers [19], so they devised a SRC steered discriminative projection (SRC-DP) [20]. By using SRC in the projected space, SRC-DP achieved better performances in face recognition [20] than the existing SRC-related feature extraction algorithms.

The success of SRC-DP proved the conclusion, namely "dimensionality reduction algorithms should be designed according to the classifiers". However, SRC needs to solve l_1 minimization problems which are time consuming. Hence, the computation cost of SRC-related feature extraction algorithms are also expensive. Recently, Zhang et al. claimed that the collaborative representation strategy played a more important role than the sparsity constraints in SRC, and suggested a collaborative representation based classifier (CRC) [21]. Compared to SRC, CRC can achieve the competitive experimental results with much less computation burden. Xu et al. used a hierarchical CRC method for face recognition [22] and confirmed that "l_2 approach to the face recognition problem is not only significantly more accurate than the state-of-the art approach, it is also more robust, and much faster" [23]. Hence, feature extraction methods based on CRC are preferred. Yang et al. proposed a collaborative representation based projection (CRP) [24] under the methodology of SPP. However, CRP is not designed according to the classification rules of CRC, and the authors even did not use CRC in the subspace obtained by CRP for classification to verify the effectiveness of CRP [24]. Hence, it cannot be guaranteed that CRP is a satisfactory feature extraction algorithm for CRC.

In this paper, we propose a feature extraction algorithm termed kernel locality-constrained collaborative representation based discriminant analysis (KLCR-DA). KLCR-DA is designed according to the classification rules of our proposed kernel locality-constrained collaborative representation based classifier (KLCRC). KLCRC is a nonlinear extension of CRC, moreover, it uses the local structures of data sets. Our reasons for devising KLCRC can be summarized as follows: (1) high-dimensional data sets, such as face image databases, usually have nonlinear structures. As a linear algorithm, CRC may be insufficient for high-dimensional data recognition. Fortunately, by using kernel approaches [7], the linear algorithms can be extended to be nonlinear ones; (2) the collaborative representation strategy means that a query image is reconstructed by using nonlocal samples [21]. However, a given image and its local samples tend to get same labels. Hence, in collaborative representation strategy, the local samples of a query image should be chosen as the reconstruction bases with higher probability than that of the nonlocal samples.

Furthermore, the goal of the proposed KLCR-DA algorithm is to seek a transformed low dimensional space in which the within-class reconstruction residual of a given data set is as small as possible and the between-class reconstruction residual is as large as possible. Hence, KLCRC can achieve better performances in the projected spaces obtained by KLCR-DA. In other words, KLCR-DA fits KLCRC well. In addition, we will show that KLCR-DA and KLCRC depend on each other. It means that the reconstruction coefficient vectors required in KLCR-DA need to be computed by KLCRC and the projection matrix required in KLCRC needs to be computed by KLCR-DA. Therefore, we use an iterative approach to obtain the optimal solution of KLCR-DA. We also prove the convergence of the iterative version of KLCR-DA. Finally, the extensive experiments on face image databases (AR database, the extended Yale B database and FERET database) and palm print database (HK PolyU) show that KLCR-DA is superior to the related algorithms.

The rest paper is organized as follows: Section 2 reviews the CRC method. The kernel locality-constrained collaborative representation based classifier (KLCRC) is proposed in Section 3. In Section 4, we design a kernel locality-constrained collaborative representation based feature extraction algorithm (KLCR-DA) which fits KLCRC well. And we present an iterative approach to obtain the optimal solution for KLCR-DA. The convergence of KLCR-DA is proved in this section. We present the comparisons between KLCR-DA and the related algorithm in Section 5. Comparative experiments for the proposed algorithms and the related algorithms are shown in Section 6. Section 7 gives the conclusions.

2. Collaborative representation based classifier (CRC)

2.1. Collaborative representation based classifier (CRC)

As we mentioned in Section 1, SRC has shown excellent performances in face recognition tasks. However, SRC needs to solve l_1-minimization problems. Hence, its computation cost is usually expensive. Compared with SRC, CRC can obtain competitive results with much less computation [21]. Hence, let us introduce CRC method formally.

Suppose that we have C classes of subjects. Denote $\mathbf{X}^i \in R^{D\times n_i}$ as the dataset of the ith class, and each column of $\mathbf{X}^i$ is a sample of class i. The entire training set is defined as $\mathbf{X} = [\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^C] \in R^{D\times n}$, where D is the dimension of samples and $n = \sum_i n_i$. Once a query image $\mathbf{y}$ comes, we represent $\mathbf{y}$ by using the training sample, namely $\mathbf{y} = \mathbf{X}\mathbf{a}$, where $\mathbf{a} = [\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^c]$ and $\mathbf{a}^i$ is the reconstruction representation vector associated with class i. If $\mathbf{y}$ is from the ith class, usually $\mathbf{y} = \mathbf{X}^i\mathbf{a}^i$ holds well. This implies that most coefficients in $\mathbf{a}^j (j \neq i)$ are nearly zeros and only $\mathbf{a}^i$ has significant nonzero entries. That means the sparse non-zero entries in $\mathbf{a}$ can faithfully reconstruct $\mathbf{y}$. Then the optimal $\mathbf{a}$ can be sought by solving the following optimization problem:

$$\mathbf{a} = \arg\min \|\mathbf{a}\|_0 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{X}\mathbf{a}. \tag{1}$$

where $\|\cdot\|_0$ is the l_0-norm, which counts the number of nonzero entries in a vector.

Solving l_0 optimization problem (1), however, is NP hard. Fortunately, recent research efforts reveal that for certain dictionaries, if the solution $\mathbf{a}$ is sparse enough, finding the solution of the l_0 optimization problem is equivalent to a l_1 optimization problem [25]. Hence, in SRC, the following l_1-minimization problem is used instead of the l_0-minimization problem:

$$\mathbf{a} = \arg\min \|\mathbf{a}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{X}\mathbf{a}. \tag{2}$$

In practice, problem (2) is always transferred to a regularization problem, which is also called Lasso [26]:

$$\min_{\mathbf{a}} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_1, \tag{3}$$

where λ is a regularization coefficient. Solving the l_1-minimization problem in Eq. (3) is still time consuming. Recently, Zhang et al. pointed out that the use of collaborative representation was more crucial than the l_1-sparsity of $\mathbf{a}$ [21]. And the l_2-norm regularization on $\mathbf{a}$ can do a similar job to l_1-norm regularization but with much less computational cost. So they proposed the collaborative representation based classifier (CRC) method which tried to solve the following problem:

$$\min_{\mathbf{a}} \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_2^2. \tag{4}$$

By taking the derivative of Eq. (4) with respect to $\mathbf{a}$ and setting it to zero, we can obtain the optimal solution $\mathbf{a} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$ [21]. λ is usually set to be a small positive real number, in our experiments $\lambda = 0.001$. Once $\mathbf{a}$ is obtained, CRC computes the residuals $r^i(\mathbf{y}) = \|\mathbf{y} - \mathbf{X}^i\mathbf{a}^i\|_2 / \|\mathbf{a}^i\|_2$ for each class i. If $r^l = \min\{r^i(\mathbf{y}) | i = 1, 2, \ldots, C\}$, $\mathbf{y}$ is assigned to class l.