Neurocomputing 99 (2013) 592–602

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Letters

Enhancing sparsity via ℓ^p $(0<p<1)$ minimization for robust face recognition

Song Guo*, Zhan Wang, Qiuqi Ruan

Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China

ARTICLE INFO

Article history:
Received 15 December 2011
Received in revised form
6 May 2012
Accepted 7 May 2012
Communicated by Ran He
Available online 25 August 2012

Keywords:
Sparse representation
Face recognition
ℓ^p minimization
ℓ^1 minimization
Sparsity ratio

ABSTRACT

Sparse representation has received an increasing amount of interest in recent years. By representing the testing image as a sparse linear combination of the training samples, sparse representation based classification (SRC) has been successfully applied in face recognition. In SRC, the ℓ^1 minimization instead of the ℓ^0 minimization is used to seek for the sparse solution for its computational convenience and efficiency. However, ℓ^1 minimization does not always yield sufficiently sparse solution in many practical applications. In this paper, we propose a novel SRC method, namely the ℓ^p $(0<p<1)$ sparse representation based classification (ℓ^p-SRC), to seek for the optimal sparse representation of a testing image. In ℓ^p-SRC, the ℓ^p $(0<p<1)$ minimization is adopted as an alternative to ℓ^0 minimization, the solution of which is sparser than that of ℓ^1 minimization used in traditional SRC. Furthermore, an iterative algorithm is introduced to efficiently solve the ℓ^p minimization problem in this paper. The extensive experimental results on publicly available face databases demonstrate the effectiveness of ℓ^p-SRC for robust face recognition.

1. Introduction

Sparsity has been a guiding principle in neuroscience, information theory and signal processing over the past few decades [1–5]. Recently, sparse representation (or coding), which encodes a natural image using only a small number of atoms parsimoniously chosen from an overcomplete dictionary, has been developed and employed in computer vision and pattern recognition area with promising results [6–8]. The finding that the firing of the neurons with respect to a given input image is typically highly sparse if these neurons are viewed as an overcomplete dictionary of base signal elements at each visual stage [1,2] provides a physiological basis for sparse representation.

Face recognition (FR) is one of the most attracting and challenging tasks in computer vision and pattern recognition owing to its applications in military, commercial and public security [9]. Various methods, such as Eigenfaces [10], Fisherfaces [11] and Laplacianfaces [12], have been proposed for FR. Recently, Wright et al. [6] reviewed the application of sparse representation in statistical signal processing community and proposed the sparse representation based classification (SRC) for robust FR.

The basic idea of SRC is to represent a test sample as a linear combination of all training samples, and classify the test sample by evaluating which class of training samples could result in the minimal reconstruction error with the associated coding coefficients. The sparse nonzero coefficients are supposed to concentrate on the training samples which are from the same class as the test sample. Naturally, the sparsest solution can be sought by solving the following optimization problem

$$\min_x \|x\|_0 \quad \text{subject to} \quad Ax=y \tag{1}$$

where y is a test sample, A is the coding dictionary with all training samples, x is the coding coefficient vector of y over A, $\|\bullet\|_0$ denotes the ℓ^0-norm, which counts the number of nonzero entries in a vector. However, the ℓ^0 minimization, which is a combinatorial optimization problem, is NP-hard and difficult even to approximate [13]. Alternatively, the ℓ^1 minimization, which is a convex problem and can be solved in polynomial time, is employed in sparse representation. Accordingly, the optimization problem of Eq. (1) can be reformulated as

$$\min_x \|x\|_1 \quad \text{subject to} \quad Ax=y \tag{2}$$

It has been revealed that the minimizations of ℓ^0-norm and ℓ^1-norm are equivalent if the solution is sufficiently sparse [4,14–15].

The use of ℓ^1-norm regularization in SRC is remarkably widespread; however, whether it is good enough to essentially characterize the signal is still an open issue. More recently, many works have been done to address this issue. By incorporating the nonnegative constraint into the sparse coefficient, Liu et al. [16] proposed a novel sparse nonnegative image representation method. Zheng et al. [17] introduced a graph Laplacian regularizer into the traditional sparse representation objective function to capture the intrinsic geometrical information of the data.

* Corresponding author.
E-mail addresses: 07112071@bjtu.edu.cn (S. Guo), 04121575@bjtu.edu.cn (Z. Wang), qqruan@center.njtu.edu.cn (Q. Ruan).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.05.028

Zeng et al. [18] proposed a novel kernelized classification framework based on sparse representation, which is performed in inner product space rather than Euclidean space.

The previously mentioned methods develop the sparse representation from different perspectives. However, to the best of our knowledge, little work has been done to enhance the sparsity of coding coefficients in sparse representation [21–22]. Based on the sparse assumption, the ℓ^1 minimization is used instead of the ℓ^0 minimization in SRC [6], but the solutions of the ℓ^1 minimization are often less sparse than those of the ℓ^0 minimization in many practical applications [24]. Furthermore, the sparser the coding coefficients are, the easier will it be to accurately determine the identity of the test sample [6]. This raises the question of whether we can find a different alternative to ℓ^0 minimization which not only find a sparser solution than ℓ^1 minimization but also be easier to be solved than ℓ^0 minimization.

Recently, ℓ^p-norm, where $0<p<1$ (in which case $\|\bullet\|_p$ is not actually a norm, though $\|\bullet\|_p^p$ satisfies the triangle inequality and induces a metric), has been proposed as a natural replacement of ℓ^1-norm for sparse signal recovery problem [19–21]. Numerical experiments in [19] showed that ℓ^p minimization with $0<p<1$ recovers sparse signals from fewer linear measurements than ℓ^1 minimization. Chartrand and Staneva [20] further generalized the result in [19] to an ℓ^p variant of the restricted isometry property. Foucart and Lai [21] presented a condition on the matrix of an underdetermined linear system which guarantees that the solution of the system with minimal ℓ^p-quasinorm is also the sparsest one.

Inspired by the recent progress in sparse representation and ℓ^p minimization, we propose a novel version of SRC via ℓ^p minimization. We denote our method by ℓ^p-SRC $(0<p<1)$ to distinguish it from the original SRC which is denoted by ℓ^1-SRC in this paper. In ℓ^p-SRC, we use ℓ^p minimization as an alternative to ℓ^0 minimization instead of using ℓ^1 minimization in ℓ^1-SRC. In this way, we can find a sparser and more accurate solution than ℓ^1-SRC does, and the optimization problem of ℓ^p minimization is much easier to be solved than that of ℓ^0 minimization.

The rest of this paper is organized as follows. We first briefly review the related works on sparse representation based classification in Section 2, and then we present a numerical example to show the superiority of ℓ^p $(0<p<1)$ minimization over ℓ^1 minimization in Section 3. In Section 4, an iteration reweighted algorithm is introduced to efficiently solve the ℓ^p minimization problem and the framework of ℓ^p-SRC is also proposed in that section. In Section 5, we evaluate the efficiency of the proposed algorithm for robust face recognition on publicly available database. Finally, we conclude the paper in Section 6.

2. Sparse representation based classification

Suppose we have n training samples from k object classes, the entire training set can be denoted by $A=[A_1,A_2,\ldots A_k]\in\mathbb{R}^{m\times n}$, where m is the dimension of sample, $A_i, i=1,2,\ldots k$ is the set of training samples of the ith object class. Given a test sample $y\in\mathbb{R}^m$, the linear presentation of y can be written in terms of all training samples A as [6]

$$y=Ax \tag{3}$$

where x is the coefficient vector. As the entries of the vector x encode the identity of the test sample y, it is tempting to obtain the solution of the linear system in Eq. (3). However, in face recognition, the linear system $y=Ax$ is typically underdetermined, so its solution is not unique. Fortunately, the test sample y can be sufficiently represented using only the training samples from the same class, so it is natural to seek for the sparsest solution. Intuitively, the sparsity of the coefficients vector can be measured by the ℓ^0-norm of it. However, it is an NP-hard problem to find the optimal solution of ℓ^0 minimization. As an alternative, the ℓ^1 minimization is employed to seek for the sparse solution for its computational convenience and efficiency, which can be expressed in Eq. (2).

After obtaining x, we can design a sparse representation based classification (SRC) in terms of the class reconstruction residual. Specifically, for each class i, let $\delta_i:\mathbb{R}^n\to\mathbb{R}^n$ be the characteristic function which select the coefficients associated with the ith class. For $x\in\mathbb{R}^n$, $\delta_i\in\mathbb{R}^n$ is a new vector whose only nonzero entries are the entries in x that are associated with class i. We can approximate the given test sample y as $\hat{y}_i=A\delta_i(x)$ using the vector δ_i from each class. The corresponding reconstruction residual for class i is defined as

$$r_i(y)=\|y-\hat{y}_i\|_2 \tag{4}$$

The identity of y is then assigned to class i who has the minimal reconstruction residual, i.e. $\min_i r_i(y)$.

In practical face recognition scenarios, the test sample y could be partially corrupted or occluded. In this case, the linear representation of y can be rewritten as

$$y=y_0+e_0=Ax_0+e_0=[A,A_e]\begin{bmatrix}x_0\\ e_0\end{bmatrix}=B\omega \tag{5}$$

where $B=[A,A_e]\in\mathbb{R}^{m\times(n+n_e)}$, A_e is the occlusion dictionary, which can be set as an orthogonal matrix, such as identity matrix, Fourier basis, Haar basis [6]. The linear system of $y=B\omega$ is also underdetermined which does not have a unique solution for ω. However, we should note that y_0 and the error e_0 have sparse representations over the training samples dictionary A and the occlusion dictionary A_e, respectively. Accordingly, the sparsest solution of ω can be obtained by solving the following extended ℓ^1 minimization problem

$$\min_\omega \|\omega\|_1 \quad \text{subject to} \quad B\omega=y \tag{6}$$

Once the sparse solution $\omega=(x/e)$ is computed, the reconstruction residual of class i can be rewritten as

$$r_i(y)=\|y-A_e e-A\delta_i(x)\|_2 \tag{7}$$

The identity of y is then assigned to class i by $\min_i r_i(y)$.

3. Sparse solution via ℓ^p minimization

The ℓ^1 minimization is widely employed as an alternative to ℓ^0 minimization in SRC. However, it is not always the case that the optimization problem of ℓ^1 minimization can find the sparsest solution. Recently, ℓ^p-norm $(0<p<1)$ has been proposed as an alternative to ℓ^0-norm for sparse signal recovery. Numerical experiments in [19–23] showed that the sparest solutions of underdetermined linear system can be obtained via ℓ^p minimization from different aspects.

For the sake of illustration, we consider the following simple numerical example, which is presented in [22] to demonstrate the superiority of weighted ℓ^1 minimization over ℓ^1 minimization. Consider the dictionary matrix

$$A=\begin{bmatrix}1&2&1\\2&1&1\end{bmatrix}\in\mathbb{R}^{2\times3}$$

and the test sample $y=\begin{bmatrix}1&1\end{bmatrix}^T\in\mathbb{R}^2$, we wish to find the correct sparse representation vector $x\in\mathbb{R}^3$ from $y=Ax$. If we use ℓ^0 minimization in Eq. (1) to seek for the optimal solution, the sparsest coding coefficients vector will be $x_0=\begin{bmatrix}0&0&1\end{bmatrix}^T$. In this case, the ℓ^0-norm of x_0 is 1, and x_0 is the correct and sparsest representation of y. Fig. 1(a) shows the set of points $x\in\mathbb{R}^3$ obeying $y=Ax$, and the ℓ^1 ball of radius 1 (the ℓ^1-norm of x_0) centered at the origin. If we use ℓ^1 minimization instead of ℓ^0 minimization to seek for the