J. Vis. Commun. Image R. 28 (2015) 15–20

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

L_0-norm sparse representation based on modified genetic algorithm for face recognition ☆

Zizhu Fan [a,*], Ming Ni [a], Qi Zhu [b], Chengli Sun [c], Lipan Kang [a]

[a] School of Basic Science, East China Jiaotong University, Nanchang, China
[b] Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[c] School of Information Engineering, Nanchang Hangkong University, Nanchang, China

ARTICLE INFO

Article history:
Received 27 May 2014
Accepted 31 December 2014
Available online 7 January 2015

Keywords:
Sparse representation
Classification
Genetic algorithm
Face recognition
L_0-norm minimization
L_1-norm minimization
Modified GA (MGA)
Optimization methods

ABSTRACT

The typical sparse representation for classification (SRC) exploits the training samples to represent the test samples, and classifies the test samples based on the representation results. SRC is essentially an L_0-norm minimization problem which can theoretically yield the sparsest representation and lead to the promising classification performance. We know that it is difficult to directly resolve L_0-norm minimization problem by applying usual optimization method. To effectively address this problem, we propose the L_0-norm based SRC by exploiting a modified genetic algorithm (GA), termed GASRC, in this paper. The basic idea of GASRC is that it modifies the traditional genetic algorithm and then uses the modified GA (MGA) to select a part of the training samples to represent a test sample. Compared with the conventional SRC based on L_1-norm optimization, GASRC can achieve better classification performance. Experiments on several popular real-world databases show the good classification effectiveness of our approach.

1. Introduction

Sparse representation based classification (SRC) [1,2] has attracted much attention in recent years. The basic idea of the typical SRC is to represent a test sample as a sparse linear combination of the training samples, and then classify the test samples based on the representation results. SRC assigns the test sample to the class which leads to the minimum representation error [3,4]. It has been widely used for many applications such as face recognition. In order to perform undersampled face recognition, in which there are very few, or even a single, training face images per subject, Deng et al. proposed an Extended SRC (ESRC) by exploiting intra-class variant dictionary [5]. Wagner et al. implemented a face recognition system based on the sparse representation to simultaneously deal with variations in illumination, image misalignment, and partial occlusion in the test image [6]. This system can work well under a variety of realistic conditions. Cheng proposed a directed L_1-graph for class analysis and classification [7]. In order to capture the nonlinear information within the data, kernel sparse representation algorithms for image classification and face recognition are proposed [8–10]. All the methods mentioned above are based on L_1-norm. Nevertheless, a number of good representation methods that are based on L_2-norm [3,11] were proposed in recent years. Table 1 concisely describes the typical representation methods based on L_1 or L_2 norm and our proposed method. In this table, we give the objective function and constrains of the problem in each method, and the features that are used in these methods.

Essentially, SRC is an L_0-norm (the number of nonzero entries in a vector) optimization problem which can theoretically yield the sparsest representation and lead to the promising classification performance [2,12]. We know that it is difficult to directly resolve it by applying usual optimization method since the L_0-norm minimization is an NP-hard problem. In pattern recognition and machine learning community, researchers often exploit L_1-norm minimization to approximate the L_0-norm minimization under the condition that the solution is sufficiently sparse [20]. This condition, however, does not always hold, particularly when the data dimensionality is very high and the data scale is relatively small. For example, in face recognition, the face image is usually very high-dimensional (the dimensionality of a small face image of size 64×64 is 4096) and the number of the training samples per subject is generally not larger than some value, say 20, even a single, in some applications, which is very small compared with the dimensionality of the face images. In this case, applying L_1-norm minimization might not achieve sufficiently sparse solution. As a

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

* Corresponding author.
E-mail address: zzfan3@163.com (Z. Fan).

http://dx.doi.org/10.1016/j.jvcir.2015.01.001
1047-3203/

result, it is not very suitable for using L_1-norm minimization to approximate the L_0-norm minimization.

In theory, the L_0-norm minimization can be effectively solved by the stochastic optimization methods such as the genetic algorithm (GA) which has been successfully applied to feature selection and other applications [21,22]. We know that the conventional GA method does not exploit the prior knowledge, e.g., the distance information of the samples. Its effectiveness and efficiency can be further improved if we employ an appropriate distance measure to evaluate the similarities between the samples in the classification. To this end, we propose the L_0-norm based SRC by exploiting a modified genetic algorithm (MGA) to select the most representative training samples which lead to the minimal representation error for a test sample in this paper. We refer to the proposed approach as GASRC. Note that the typical SRC assumes that the samples from a single class lie on a subspace [2]. This implies that a sample can be represented by the other samples from the class to which this sample belongs. On the other hand, a sample and its neighbors generally belong to the same class, and these neighbors have a close correlation with this sample. They play more important role in representing the sample than other samples from the same class. Accordingly, it is natural that GASRC is based on the following prior knowledge. When representing a test sample, the neighbors of this test sample are more representative than those that are not its neighbors [11,23]. In other words, using the neighbors of a test sample to represent it can yield the less representation error, compared with using those samples that are not neighbors of this test sample.

In this work, we pay attention on the classification performance of GASRC. It aims to select the most representative training samples, termed the representation set which can yield the sparsest representation and lead to promising classification performance, to represent the test sample by applying MGA approach. As discussed above, the representation set of a test sample tends to contain its neighbors or the training samples that are more similar to the test sample. GASRC divides the representation set of a test sample into two parts. The first part contains some appropriate nearest neighbors of this test sample. The second part consists of the other representative training samples obtained by MGA. When selecting the second part of the representation set for a test sample, MGA needs to combine the determined neighbors of the test sample. If a number of training samples including the neighbors of the test sample yield the minimal representation error, MGA determines them as the representation set of the test sample. Finally, GASRC uses the determined representation set to represent the test sample and classify it.

Our work has the following contributions.

First, GASRC is suitable to be directly applied in the learning settings in which the data are high-dimensional and the number of the training data is small. It is worth noting that if the traditional SRC based on L_1-norm optimization deals with very high-dimensional data when the scale of the training set is relatively small, it often needs the dimensionality reduction such as PCA [14] to reduce the data dimensionality. Otherwise, the traditional SRC tends to yield the non-sparse solution, which might increase its classification error rate.

Second, GA is applied in the sparse representation based classification for the first time in pattern recognition community, to our best knowledge. It can be used to effectively select the representation set for a test sample. MGA directly exploits the nearest neighbors of the test sample as a part of the representation set. This scheme can effectively reduce the search state space and enhance the selection efficiency. Moreover, this scheme guarantees to obtain the representation set that contains the neighbors of the test sample or the samples that have close correlation with the test sample.

Third, besides the traditional SRC approach, our proposed algorithm can be generally applied to the other representation based approaches, as long as these approaches essentially exploit the L_0-norm optimization. Compared with the conventional SRC based on L_1-norm, the proposed GASRC algorithm can achieve better classification performance, particularly when the data are high-dimensional and the number of the training samples is small. Experiments on several popular real-world databases justify the good classification effectiveness of our approach.

The rest of this paper is organized as follows. Section 2 gives a review of the traditional SRC based on L_0-norm and L_1-norm optimization. Section 3 presents our proposed approach, i.e., GASRC. Section 4 reports the experiment results on the popular high-dimensional image data sets. And we offer the conclusions in Section 5.

2. SRC

Assume that there are N training samples $X = [x_1, x_2, \ldots, x_N] \in R^{m\times N}$ where $x_i \in R^m$ $(i = 1, 2, .., N)$ denotes the ith training sample. Given a test sample $y \in R^m$, the traditional SRC essentially seeks the sparsest solution to $y = X\beta$, where $\beta = [\beta_1, \beta_2, \ldots, \beta_N]^T$ and β_i is the representation coefficient associated with the ith training sample, by solving the following optimization problem [2]

$$\hat{\beta}_0 = \arg\min \|\beta\|_0 \quad \text{subject to} \quad y = X\beta, \tag{1}$$

where $\|\beta\|_0$ denotes the L_0-norm of the representation coefficient vector β, which is the number of nonzero entries in β. Often, the optimization problem in Eq. (1) is approximately solved by using the following L_1-norm optimization,

$$\hat{\beta}_1 = \arg\min \|\beta\|_1 \quad \text{subject to} \quad y = X\beta, \tag{2}$$

where $\|\beta\|_1 = \sum_{i=1}^{N} |\beta_i|$. It is worth noting that the L_1-norm optimization in Eq. (2) works well if the solution is sufficiently sparse. Otherwise, as discussed in Section 1, the L_1-norm optimization may not

Table 1
The description of the typical representation methods and our method (GASRC).

Methods	Objective function	Constrains	Features used
CRC[3]	$\min_\alpha\{\|\alpha\|_2^2\}$	$y = X\alpha$ or $\|y - X\alpha\|_2^2 \leqslant \delta$	PCA, down sample
SRC[2]	$\min_\alpha\{\|\alpha\|_1\}$	$y = X\alpha$ or $\|y - X\alpha\|_2^2 \leqslant \delta$	PCA, LDA, down sample, random faces
WSRC [13]	$\min_\alpha\{\|w\alpha\|_1\}$	$y = X\alpha$ or $\|y - X\alpha\|_2^2 \leqslant \delta$	PCA, LDA, down sample, random faces
ESRC[5]	$\min_{\alpha,\beta}\{\|\begin{matrix}\alpha\\ \beta\end{matrix}\|_1\}$	$y = X\alpha + D_I\beta$ or $\|y - [X, D_I][\begin{matrix}\alpha\\ \beta\end{matrix}]\|_2^2 \leqslant \delta$	Gabor features, down sample, random faces
KSRC [8]	$\min_\alpha\{\|\alpha\|_1\}$	$\varphi(y) = \varphi(X)\alpha$ or $\|\varphi(y) - \varphi(X)\alpha\|_2^2 \leqslant \delta$	KPCA, KFDA, random projection
GASRC	$\min_\alpha\{\|\alpha\|_0\}$	$y = X\alpha$ or $\|y - X\alpha\|_2^2 \leqslant \delta$	PCA, original images

PCA: principal component analysis [14,15], LDA: linear discriminant analysis [16,17], KPCA: kernel principal component analysis [18], and KFDA: kernel Fisher discriminant analysis [19].