Random subspace based ensemble sparse representation

CrossMark

Jing Gu*, Licheng Jiao, Fang Liu, Shuyuan Yang, Rongfang Wang, Puhua Chen, Yuanhao Cui, Junhu Xie, Yake Zhang

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University, Xi'an 710071, China

ARTICLE INFO

ABSTRACT

In this paper, a new random subspace based ensemble sparse representation (RS_ESR) algorithm is proposed, where the random subspace is introduced into sparse representation model. For high-dimensional data, the random subspace method can not only reduce dimension of data but also make full use of effective information of data. It is not like traditional dimensionality reduction methods that may lose some information of original data. Additionally, a joint sparse representation model is emloyed to obtain the sparse representation of a sample set in the low dimensional random subspace. Then the sparse representations in multiple random subspaces are integrated as an ensemble sparse representation. Moreover, the obtained RS_ESR is applied in classical clustering and semi-supervised classification. The experimental results on different real-world data sets show the superiority of RS_ESR over traditional methods.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

A large amount of data is easily obtained with the ever-accelerated updating of information technology. As an effective technique for analyzing the sparsity of large data, sparse representation (SR) [1-2] method emerges as the times require. In recent years, SR has attracted attention of many researchers and been successfully applied in image classification and other fields, such as signal reconstruction [3], image super-resolution [4], visual tracking [5], face recognition [6–7], etc. At first, Wright et al. [8] proposed a sparse model-based method for facial images classification. During the same time period, a sparse subspace clustering algorithm was presented by Elhamifar et al. [9], which combines the SR technique with spectral clustering to segment different moving objects in the video. Hereafter, many improved sparse representation based classification methods keep emerging [10-17].

Although sparse representation based classification methods are very effective, low memory problem occurs when sparse representation is used to deal with high dimensional data. Since sparse representation method updates the similarity between every two samples in each iteration, it needs massive computation and storage, especially for the high dimensional dataset. In fact, the data usually lies in a high dimensional space in many real applications. It is well known that an effective method to handle the high dimensional data is dimensionality reduction. The common dimen-sionality reduction approaches include principal component analysis [18], linear discriminant analysis [19], locality preserving projections [20], etc. After the rise of SR, a new dimensionality reduction method called sparsity preserving projections (SPP) was proposed [21]. But whatever dimensionality reduction approach is used, the reduction of the dimension of data in the original space results in information loss. In this case, the spatial relationship among the samples in lower dimensional space may be changed, which affects the following clustering or classification results.

To make full use of potential information, a random subspace method is used in this paper. The main difference between it with the traditional dimensionality reduction approaches is that the random subspace method randomly samples many lower-dimensional subspaces from the original high-dimensional space. The random subspace method has been successfully applied in classifier ensemble [22–23], which is more robust to noise and redundant information than a single classifier. This indicates that several random lower-dimensional subspaces include more effective information than the original high-dimensional space. The success of the random subspace method in classifier ensemble motivates us to apply it to reduce dimension.

In this paper, we combine the SR with the random subspace, and propose a new algorithm called random subspace based ensemble sparse representation (RS_ESR). Firstly, multiple subspaces are obtained from the original high-dimensional space by using the random subspace method. Then a joint sparse representation model is used to simultaneously calculate the sparse representation of samples in each subspace. Afterwards, the sparse repre-

* Corresponding author.
  E-mail address: xuer6126@126.com (J. Gu).

sentations in all lower-dimensional subspaces are integrated into an ensemble sparse representation. Finally, the proposed RS_ESR is applied in dataset clustering, image segmentation and semi-supervised classification.

The remaining parts of this paper are organized as follows. Section 2 reviews the classical SR model and random subspace method. The random subspace based joint SR model and RS_ESR method are described in Section 3. Section 4 presents the RS_ESR based clustering and semi-supervised classification algorithms. Experimental results in different datasets and images are contained in Section 5. Finally, Section 6 provides conclusions and discussion of possible improvements in future work.

2. Related work

2.1. Sparse representation (SR)

The fundamental of SR is that any one $\mathbf{x} \in R^d$ of a dataset can be represented by a linear combination of bases in a dictionary $D \in R^{d \times l} (d \ll l)$. The weights of all atoms in the linear combination are sparse and named as SR. The SR $\mathbf{z}$ of sample $\mathbf{x}$ can be obtained by solving the following model [1]

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \ s.t. \ \mathbf{x} = D\mathbf{z}, \tag{1}$$

where $\|\cdot\|_0$ is ℓ_0 norm of a vector. The objective function in (1) is an NP-hard problem, so it is replaced with

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \ s.t. \ \mathbf{x} = D\mathbf{z}, \tag{2}$$

where $\|\cdot\|_1$ is ℓ_1 norm of a vector.

2.2. Random subspace

The random subspace [24] method was first presented in decision forest, in which multiple decision trees were generated in multiple random subspaces and then integrated into a classifier. On that basis, random forest [25] and rotation forest [26] were consecutively put forward. The method to get the random subspace includes two steps, as detailed in Algorithm 1.

One possible problem is that multiple lower-dimensional datasets obtained by the random subspace method may not include the discriminative information of original dataset. In fact, it seldom happens. The larger q and p, the smaller the probability that no discriminative information is selected in q random subspaces is. The specific reasons have been analyzed in [24] and [27], so we will not cover them in this paper.

3. Random subspace based ensemble sparse representation (RS_ESR)

3.1. Random subspace based ensemble sparse representation

A proposed multi-task low-rank affinity pursuit (MLAP) [28] algorithm integrated multiple types of features and effectively used cross-feature information of multiple features. Inspired by the MLAP algorithm, we consolidate multiple random subspace sets in a similar way. But a great deal of computer's internal storage and computer time are spent to solve the nuclear norm of the matrix in MLAP. Thus, to improve the efficiency of algorithm, we introduce a novel random subspace based joint sparse representation model

$$\min_{Z^1,\cdots,Z^q} \sum_{t=1}^{q} \left(\|Z^t\|_1 + \alpha \|S^t - S^t Z^t\|_F^2 \right) + \beta \|Z\|_{2,1}$$

$$s.t. S^t = X(\mathbf{r}^t), \ diag(Z^t) = 0, \tag{4}$$

where X denotes the data set in original space; S^t is tth random subspace set of X obtained by Algorithm 1; Z^t is sparse representation of S^t, and $\|Z^t\|_1 = \sum_{i,j} |Z^t_{ij}|$; $\|\cdot\|_F$ is Frobenius norm; $\alpha > 0$ and $\beta > 0$ are parameters to balance the effect of different parts; the $n^2 \times q$ matrix Z is constructed in the same way as the MLAP algorithm; $\|Z\|_{2,1} = \sum_i \sqrt{(\sum_j |Z_{ij}|)^2}$. The constraint $diag(Z^t) = 0$ is to avoid that the solution of (4) is the identity matrix.

We employ the inexact augmented Lagrange multiplier method, also called alternating direction method of multipliers (ADMM) [29] to solve (4). Firstly, we introduce two auxiliary variables K^t and L^t, so the objective function in (4) is converted into equivalent form as follows:

$$\min_{\substack{K^1,\cdots,K^q \\ L^1,\cdots,L^q \\ Z^1,\cdots,Z^q}} \sum_{t=1}^{q} \left(\|K^t\|_1 + \alpha \|S^t - S^t L^t\|_F^2 \right) + \beta \|Z\|_{2,1}$$

$$s.t. \ Z^t = K^t, \ diag(K^t) = 0, Z^t = L^t$$

Then minimize the following augmented Lagrange function

$$\beta \|Z\|_{2,1} + \sum_{t=1}^{q} \left(\|K^t\|_1 + \alpha \|S^t - S^t L^t\|_F^2 \right)$$

$$+ \sum_{t=1}^{q} \left(\langle U^t, Z^t - K^t \rangle + \langle V^t, Z^t - L^t \rangle + \frac{\mu}{2} \|Z^t - K^t\|_F^2 + \frac{\mu}{2} \|Z^t - L^t\|_F^2 \right)$$

where $U^1, \cdots, U^q$ and $V^1, \cdots, V^q$ are Lagrange multipliers and $\mu > 0$ is a penalty parameter. By the ADMM method, the problem (4) is divided into several sub-problems which have closed-form solutions. The solution procedure of (4) is detailed in Algorithm 2.

In Algorithm 2, we utilize the soft thresholding operator [30] to solve (5) and the objective function in (6) can be solved according to Lemma 3.2 in [31].

Let $\tilde{Z}^1, \cdots, \tilde{Z}^t, \cdots, \tilde{Z}^q$ be the optimal solution of the problem (4). These sparse representations $\tilde{Z}^1, \cdots, \tilde{Z}^t, \cdots, \tilde{Z}^q$ in random subspaces are integrated by

$$E_{ii'} = \sqrt{\sum_{t=1}^{q} \left(\tilde{Z}^t_{ii'} \right)^2}, \tag{7}$$

where $\tilde{Z}^t_{ii'}$ $(1 \leq i \leq n, 1 \leq i' \leq n)$ and $E_{ii'}$ are the i'th element in ith row of $\tilde{Z}^t$ and E respectively, and E is viewed as ensemble sparse representation.

From the methodology point of view, the obtained ensemble sparse representation can be directly used for existing clustering and classification methods based on sparse representation. On the other side, by integrating the multiple sparse representations in subspaces, effective information in multiple lower-dimensional datasets is reinforced and redundant information is weakened. So the integrated ensemble sparse representation is more helpful for the clustering and classification than the spare representations in subspaces. Algorithm 3 summarizes the generation process of the ensemble sparse representation E. It reflects the similarity among samples in the original dataset X. The larger the ensemble sparse representation coefficient $E_{ii'}$ is, the more similar the corresponding ith and i'th samples are, and vice versa.

3.2. Computational complexity analysis

The computational complexity of RS_ESR is analyzed in this subsection. We suppose that the number of iterations is e. The time complexity of the proposed RS_ESR algorithm is $O(n^2 pqe)$. In fact, if the random subspace method is not used for RS_ESR, the RS_ESR algorithm is equivalent to the existing L1-graph [32]. The computing complexity of L1-graph is $O(n^2 de)$. Generally, the initial dimension of sample d is close to pq. So the time complexities of RS_ESR and L1 are roughly equivalent. Furthermore, since