ARTICLE IN PRESS

Neurocomputing xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Neurocomputing

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

Discriminative analysis-synthesis dictionary learning for image classification

Meng Yang[a,b,*], Heyou Chang[c], Weixin Luo[a,b]

[a] School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China
[b] College of Computer Science & Software Engineering, Shenzhen University, Shenzhen, China
[c] Nanjing University of Science and Technology, Nanjing, China

ARTICLE INFO

Communicated by Xinmei Tian

Keywords:
Discriminative
Analysis-synthesis dictionary learning
Image classification

ABSTRACT

Dictionary learning has played an important role in the success of sparse representation. Although discriminative synthesis dictionary learning for sparse representation with a high-computational-complexity l_0 or l_1 norm constraint has been well studied for image classification, jointly and discriminatively learning an analysis dictionary and a synthesis dictionary is still in its infant stage. As a dual of synthesis dictionary, the recently developed analysis dictionary can provide a complementary view of data representation, which can have a much lower time complexity than sparse synthesis representation. Although several class-specific analysis-synthesis dictionary, which may have a big correlation between different classes' dictionaries, have been developed, how to learn a more compact and discriminative universal analysis-synthesis dictionary is still open. In this paper, to provide a more complete view of discriminative data representation, we propose a novel model of discriminative analysis-synthesis dictionary learning (DASDL), in which a linear classifier based on the coding coefficient is jointly learned with the dictionary pair, thus the performance of the classifier and the representational power of the dictionary pair being considered at the same time by the same optimization procedure. The size of the learned dictionaries can be very small since the analysis-synthesis dictionary is shared by all class data. An iterative algorithm to efficiently solve the proposed DASDL is presented in this paper. The experiments on face recognition, gender classification, action recognition and image classification clearly show the superiority of the proposed DASDL.

1. Introduction

Inspired by the success of sparse coding in image processing [1,2] and the discovery of sparsity mechanism in human vision perception [3], sparse representation has been widely applied to a variety of problems, ranging from computer vision to pattern recognition [4]. For example, based on sparse representation, Zhou et al. [5] have developed a practical system for fast detection of defects occurring on the surface of bottle caps. As indicated by [6], a dictionary (i.e., a set of representation bases) plays an important role in the success of sparse representation, and learning the desired dictionary from training data instead of using off-the-shelf bases (e.g., wavelets) has led to state-of-the-art results in many practical applications, such as image denoising [1], face recognition [7,8], and image classification [9,10].

According to the way of encoding input signals, the dictionary used in sparse representation could be categorized into synthesis dictionary [11] and analysis dictionary [12]. Therefore dictionary learning approaches could be mainly divided into three categories: synthesis dictionary learning, analysis dictionary learning, and analysis-synthesis dictionary pair learning.

Synthesis dictionary learning: This kind of dictionary learning methods has been well studied and most of existing dictionary learning methods belong to this category. Traditionally, a linear combination of dictionary atoms is used to synthetically reconstruct a signal. Given the training data $\mathbf{X} \in \mathcal{R}^{m\times n}$, where n is the number of training samples and m is the dimensionality of a training sample, a synthesis dictionary $\mathbf{D}$ could be learned via

$$\min_{\mathbf{D},\mathbf{A}} \| \mathbf{X} - \mathbf{DA} \|_F^2 + \Theta(\mathbf{D}, \mathbf{A}) \quad s.\ t.\ \| \mathbf{d}_i \|_2^2 = 1 \ \ \forall\ i \tag{1}$$

where $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_k] \in \mathcal{R}^{m\times k}$ is a desired synthesis dictionary, k is the number of dictionary atoms, $\mathbf{A} \in \mathcal{R}^{k\times n}$ is the representation coefficient matrix, and $\Theta(\mathbf{D}, \mathbf{A})$ is a task-driven regularizer (e.g., sparsity of $\mathbf{A}$) [13].

A representative unsupervised dictionary learning method is K-SVD [11], which has shown promising performance in image restoration.

* Corresponding author.
E-mail addresses: yangmengpolyu@gmail.com (M. Yang), cv_hychang@126.com (H. Chang), 493752864@qq.com (W. Luo).

http://dx.doi.org/10.1016/j.neucom.2016.09.037
Received 16 October 2015; Received in revised form 15 August 2016; Accepted 18 September 2016
Available online xxxx
0925-2312/

Please cite this article as: Yang, M., Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.037

Then Smith and Elad [14] improved K-SVD dictionary learning by modifying the dictionary learning update and sparse coding. Tillmann [15] has studied the computational intractability of exact and approximate dictionary learning. Moreover, jointly learning a dictionary and a nonlinear dimension reduction has also been proposed [16]. For classification tasks, discriminative derivations [7,8] of K-SVD were developed by jointly learning a dictionary and a classifier based on the coding coefficients. The learned dictionary in [7] is a shared dictionary that could represent data from all classes. The learned synthesis dictionary can be class-specific, in which each dictionary atom has a single class label, e.g., label-consistent K-SVD (LC-KSVD) [8] associated label information with each dictionary atom. Following this line, class-specific dictionary learning, e.g., dictionary learning with structured incoherence (DLSI) [9] and Fisher discrimination dictionary learning (FDDL) [10], were developed by requiring that a class-specific sub-dictionary should properly represent the corresponding class but being weak for the other classes. To fully explore the advantage of class-specific dictionaries and shared dictionaries, some hybrid dictionary learning models (i.e., including both shared and class-specific dictionary atoms) were proposed. Zhou et al. [17] learned a hybrid dictionary with a Fisher-like regularizer on the representation coefficients, while Kong et al. [18] learned a hybrid dictionary by introducing an incoherence penalty term to the class-specific sub-dictionaries.

Although promising results have been reported in the synthesis dictionary learning, the coding coefficients need to be computed by the time-consuming sparse coding. Moreover, synthesis sparse coding cannot give an intuitive illustration like feature transformation (e.g., image operator), so analysis sparse coding, as a dual analysis viewpoint, has been studied recently.

Analysis dictionary learning: Recently, analysis dictionary learning has attracted much attention [12,19]. The representative analysis dictionary learning approach is the analysis K-SVD [12], which is an unsupervised learning approach. Compared to synthesis dictionary, analysis dictionary representation, i.e., $\mathbf{\Omega X}$, has some unique merits, e.g., generalization to feature transformation and image convolution. Rubinstein et al. [12] proposed to analytically represent the training data (here we denote an analysis dictionary by $\mathbf{\Omega} = [\mathbf{w}_1^T; \dots; \mathbf{w}_k^T] \in \mathcal{R}^{k\times m}$ to distinguish it from the synthesis dictionary $\mathbf{D}$, where each row of $\mathbf{\Omega}$ is an analysis atom):

$$\min_{\mathbf{\Omega}} \parallel \mathbf{\Omega X} \parallel_0 + \Theta(\mathbf{\Omega}) \quad s.\ t.\ \parallel \mathbf{w}_i \parallel_2^2 \le 1\ \forall\ i \tag{2}$$

where $\|\cdot\|_0$ counts the number of non-zero entries, and $\Theta(\mathbf{\Omega})$ is a task-driven regularizer (e.g., $\mathbf{\Omega}$ can not be a trivial solution). The analysis dictionary, $\mathbf{\Omega}$, aims to make $\mathbf{\Omega X}$ a sparse representation coefficient matrix, serving a similar role to $\mathbf{A}$.

Although promising result in image restoration has been reported in [12], analysis K-SVD is still computationally complicated and not designed for classification task due to the lack of enough discrimination.

Analysis-synthesis dictionary pair learning: Synthesis dictionary represents an input signal by using a linear combination of dictionary atoms, while analysis dictionary directly transforms a signal to a sparse feature space by multiplying the signal, which provides a complementary view of data representation. In the view of computation, synthesis dictionary representation involves a high-time-complexity sparse coding, while analysis-synthesis dictionary representation can reconstruct a signal with analysis coding coefficient, which can be fast computed by a linear projection. To provide a faster and more complete view of data representation, analysis-synthesis dictionary learning has been proposed. Rubinstein and Elad [20] proposed an analysis-synthesis dictionary learning (ASDL) model for the task of image processing. In ASDL, a pair of synthesis dictionary and analysis dictionary was learned from an image patch set to pursue their good representation ability and sparsity of coding coefficients. Different from ASDL, Gu et al. [21] proposed a projective dictionary pair learning (PDPL) for image classification. In PDPL, the synthesis class-specific dictionary and analysis class-specific dictionary were required to well represent the corresponding class but poorly represent the other classes. For classification tasks, obviously ASDL ignores to introduce discrimination into the learned dictionaries. Although PDPL has introduced the class-specific structure into dictionary pair learning, it learned a pair of dictionaries for each class, resulting in high redundancy between different class-specific dictionaries and large size of the learned dictionary. How to learn a more compact and discriminative universal analysis-synthesis dictionary is still open.

In this paper, we propose a novel model of discriminative analysis-synthesis dictionary learning (DASDL). In the proposed model, a classifier based on the coding coefficient is jointly learned with the dictionary pair. Incorporating the classification stage directly into the dictionary-learning procedure can gain a discriminative power for the dictionary. For the coding coefficient, a thresholding operator is used to make it sparse and a regression residual of its projection to its label space is minimized to embed discrimination. The proposed DASDL is evaluated on the application of face recognition, gender classification, action recognition and image classification. Compared with other state-of-the-art dictionary learning methods, DASDL has better or competitive performance in various classification tasks.

In our paper, the number of training samples, the number of classes, the number of dictionary atoms, and the dimensionality of signal are denoted by n, c, k and m, respectively. We denote matrices and vectors by bold-italic capital letters and lowercase letters, respectively. The rest of this paper is organized as follows. Section 2 briefly introduces the related work. Section 3 presents the proposed DASDL model. Section 4 conducts experiments, and Section 5 concludes the paper.

2. Related work

Rubinstein and Elad [20] proposed an analysis-synthesis dictionary learning (ASDL) for image deblurring. ASDL explicitly learned a pair of analysis dictionary and synthesis dictionary via

$$\min_{\mathbf{\Omega},\mathbf{D},\lambda} \parallel \mathbf{X} - \mathbf{D}S_\lambda(\mathbf{\Omega X}) \parallel_F^2 \quad s.\ t.\ \parallel \mathbf{w}_i \parallel_2 = 1\ \ \forall\ i \tag{3}$$

where $\mathbf{\Omega} = [\mathbf{w}_1^T; \dots; \mathbf{w}_k^T]$ is an analysis dictionary, $\mathbf{w}_i \in \mathcal{R}^m$, $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_k]$ is a synthesis dictionary, $\mathbf{d}_i \in \mathcal{R}^m$, $\lambda = [\lambda_1, \dots, \lambda_k]$ and λ_i is the threshold for the i-th row of $\mathbf{\Omega}$. $S_\lambda(\cdot)$ is a hard thresholding function operating on matrices with k rows, thresholding the i-row by λ_i (i.e., $\forall\ j$, if $|[\mathbf{\Omega X}]_{i,j}| \ge \lambda_i$, $[S_{\lambda_i}(\mathbf{\Omega X})]_{i,j} = [\mathbf{\Omega X}]_{i,j}$; otherwise, $[S_{\lambda_i}(\mathbf{\Omega X})]_{i,j} = 0$). The l_2 norm of $\mathbf{d}_i$ or $\mathbf{w}_i$ is usually required to be unity in order to avoid a trivial solution. It is easy to see that in ASDL, the learned dictionary pair is only required to well represent the training samples, without considering its application to classification task.

Following ASDL, Gu et al. [21] proposed a projective dictionary pair learning (PDPL) for pattern classification. The core idea of PDPL is that both analysis and synthesis dictionaries are class-specific dictionaries. Defining $\mathbf{D}_c$ and $\mathbf{\Omega}_c$ as the cth-class synthesis sub-dictionary and analysis sub-dictionary respectively, PDPL can be formulated as

$$\min_{\mathbf{\Omega},\mathbf{D}} \sum_{c=1}^{C} \parallel \mathbf{X}_c - \mathbf{D}_c\mathbf{\Omega}_c\mathbf{X}_c) \parallel_F^2 + \lambda \parallel \mathbf{\Omega}_c\overline{\mathbf{X}}_c \parallel_F^2 \tag{4}$$

where $\overline{\mathbf{X}}_c$ is the training data of all classes except class c. Eq. (4) requires the class-specific dictionary pair of class c (e.g. $\mathbf{D}_c$ and $\mathbf{\Omega}_c$) well represent the c-th training data but poorly represent the training data of $\overline{\mathbf{X}}_c$. This structured dictionary regularization in PDPL makes the learned dictionary pair discriminative for classification task, and promising results have been presented in [21].

With the theoretical development of synthesis and analysis dictionary learning, dictionary learning based sparse representation has been widely applied to image denoising [1], face recognition [7,8], image classification [9,10] and detection [5]. For example, Song et al.