Signal Processing 128 (2016) 28–39

Contents lists available at ScienceDirect

Signal Processing

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

Efficient Fisher Discrimination Dictionary Learning

Rui Jiang [a], Hong Qiao [a,b,*], Bo Zhang [c]

[a] State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] CAS Center for Excellence in Brain Science and Intelligence Technology (CEBSIT), Chinese Academy of Sciences, Shanghai 200031, China
[c] LSEC and Institute of Applied Mathematics, AMSS, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 17 July 2015
Received in revised form
15 December 2015
Accepted 22 March 2016
Available online 25 March 2016

Keywords:
Fisher discrimination dictionary learning
Nesterov's accelerated gradient method
Face recognition
Domain adaptation

ABSTRACT

Fisher Determination Dictionary Learning (FDDL) has shown to be effective in image classification. However, the Original FDDL (O-FDDL) method is time-consuming. To address this issue, a fast Simplified FDDL (S-FDDL) method was proposed. But S-FDDL ignores the role of collaborative reconstruction, thus having an unstable performance in classification tasks with unbalanced changes in different classes. This paper focuses on developing an Efficient FDDL (E-FDDL) method, which is more suitable for such classification problems. Precisely, instead of solving the original Fisher Discrimination based Sparse Representation (FDSR) problem, we propose to solve an Approximate FDSR (A-FDSR) problem whose objective function is an upper bound of that of FDSR. A-FDSR considers the role of both the discriminative reconstruction and the collaborative reconstruction. This makes E-FDDL stable when dealing with classification tasks with unbalanced changes in different classes. Furthermore, fast optimization strategies are applicable to A-FDSR, thus leading to the high efficiency of E-FDDL which can be explained by analysis on convergence rate and computational complexity. We also use E-FDDL to accelerate the Shared Domain-adapted Dictionary Learning (SDDL) algorithm which is a FDDL based new method for domain adaptation. Experimental results on face and object recognition demonstrate the stable and fast performance of E-FDDL.

1. Introduction

Fisher Discrimination Dictionary Learning (FDDL), which is an interesting variant of Sparse Representation Classifier (SRC) [1,2], was proposed recently in [3,4]. The success of FDDL can be ascribed to three key ideas. The first one is *discriminative reconstruction*. Different from those Dictionary Learning (DL) methods for learning a shared dictionary (see, e.g., [5–14]), FDDL learns a dictionary composed of class-specific sub-dictionaries. Each sub-dictionary is encouraged to well reconstruct the corresponding training examples, but poorly reconstruct the others. Thus, the class-wise reconstruction errors can be used for classification. The second one is *collaborative reconstruction*, which means that, the reconstruction of each training example should be performed collaboratively over the whole dictionary. This idea distinguishes FDDL from those DL methods for learning a dictionary for each class independently (see, e.g., [15–20]). And the third one is *discriminative representation*. This idea implies that the representation coefficients of the training examples should have a small within-class variance and a large between-class scatter. Thus, the representation coefficients can be exploited in classification. FDDL is an essentially supervised DL method. In this category, many methods have been proposed. Here, we review some state-of-the-art methods. Discriminative KSVD (D-KSVD) [12] is a supervised DL method designed for face recognition. Label Consistent KSVD (LC-KSVD) [13,14] is an extension of D-KSVD. The Shared Domain-adapted DL (SDDL) method [21,22] is a FDDL based supervised DL method for domain adaptation. For more information of this class of DL methods, see a recent survey [23].

The tradeoff among the three ideas indeed leads to a good performance of FDDL. However, the original FDDL model is complicated, and the derived DL method, i.e., O-FDDL, is often time-consuming. To address this issue, a simplified FDDL model was presented in [4], which is obtained from the original FDDL model under the assumption that each training example can only be reconstructed by columns in its corresponding sub-dictionary. The simplified model has much fewer variables, so the S-FDDL method is much faster than the O-FDDL method. Nevertheless, S-FDDL is a class-by-class DL method, which means that, in the learning process of S-FDDL, only the discriminative reconstruction and the discriminative representation are considered, but the idea of collaborative reconstruction is ignored. In [4], the equivalence of S-FDDL and O-FDDL is empirically investigated in various image classification problems. Based on the experimental results, it was

* Corresponding author at: State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.
E-mail addresses: jiangrui627@163.com (R. Jiang), hong.qiao@ia.ac.cn (H. Qiao), b.zhang@amt.ac.cn (B. Zhang).

http://dx.doi.org/10.1016/j.sigpro.2016.03.013
0165-1684/

concluded in [4] that, for the image classification problems in which data have similar variations in different classes, the classification accuracy of S-FDDL is often close to that of O-FDDL, so S-FDDL can serve as an efficient FDDL in these tasks; however, for the classification tasks involving data with unbalanced variations in different classes, such as some face recognition tasks in which the changes in illumination, accessory, expression, pose or view are often non-uniform, the classification accuracy of S-FDDL is always worse than that of O-FDDL due to the ignorance of the collaborative construction. For more details about this, see Section 6.1.1 of [4].

In this paper, we develop an Efficient FDDL (E-FDDL) method, which is particularly suitable for the classification tasks involving data with unbalanced variations in different classes. To do this, we first notice that the O-FDDL method is an iterative optimization process to alternatively solve two problems until convergence: the Fisher Discrimination based Sparse Representation (FDSR) problem and the Dictionary Update (DU) problem. The DU strategy of O-FDDL used in [3,4] is very fast, but the optimization procedure of FDSR suffers from a great deal of execution time. Our E-FDDL addresses this issue by solving an Approximate FDSR (A-FDSR) problem whose objective function is an upper bound of that of the original FDSR problem in the O-FDDL method. A-FDSR has two advantages. Firstly, A-FDSR considers all the three key ideas of FDDL. In image classification tasks involving data with unbalanced variations in different classes, this property ensures that the E-FDDL method has a better and more stable performance compared with the S-FDDL method. Secondly, A-FDSR can be split into several subproblems, and the dual problem of each subproblem is smooth, strongly convex and has fewer variables than the primal problem. This makes it possible to apply fast optimization strategies, such as Nesterov's accelerated gradient method [24], to the dual problems of these subproblems, thus effectively accelerating O-FDDL. To explain this more clearly, we analyze and compare the convergence rates and computational complexities of the key steps in solving FDSR and A-FDSR, respectively. In the experimental section, the stability and efficiency of E-FDDL are verified in face recognition tasks on two popular databases.

In addition, we evaluate the performance of the E-FDDL method in domain adaptation applications. The SDDL method, which is a FDDL based discriminative DL method, was proposed recently in [21,22] and has been proved to be effective in object recognition tasks involving data from multiple visual domains. We use our E-FDDL algorithm to replace the O-FDDL algorithm in the original SDDL method, thus obtaining a more efficient version of SDDL which is called Efficient SDDL (E-SDDL) in this paper. Object recognition experiments on two real-world databases involving four different domains were conducted to show that E-SDDL keeps the good recognition accuracy of SDDL but is much faster than SDDL. Obviously, this superiority owes much to the stability and efficiency of the proposed E-FDDL.

The remaining part of this paper is organized as follows. We give a brief review of the FDDL and FDDL-based SDDL methods in Section 2. In Section 3, we present the details of the proposed E-FDDL and E-SDDL methods. The experimental results on face and object recognition are presented in Sections 4.1 and 4.2, respectively. The final section concludes this paper.

2. A review of Fisher Discrimination Dictionary Learning

In this section, we first briefly review the original and simplified models of FDDL, and their corresponding DL methods. Then we give a review on the recently proposed domain-adaptive discriminative DL method called SDDL, which can be viewed as a modification of FDDL for domain adaptation applications.

2.1. The O-FDDL model and the O-FDDL method

Given the training examples $\boldsymbol{Y} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_C] \in \mathbb{R}^{n\times N}$, where C is the number of classes (known, fixed), n is the dimensionality of these N training examples, $\boldsymbol{Y}_j \in \mathbb{R}^{n\times N_j}$ is a matrix composed of N_j training examples with class label j. Let the desired over-complete dictionary be $\boldsymbol{D} = [\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_C] \in \mathbb{R}^{n\times K}$ with $n < K$, where K is the number of columns in the whole dictionary, $\boldsymbol{D}_j \in \mathbb{R}^{n\times K_j}$ is the sub-dictionary associated with class j and K_j is the number of columns in this sub-dictionary. We denote the sparse representation matrix of $\boldsymbol{Y}$ over $\boldsymbol{D}$ by $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_C] \in \mathbb{R}^{K\times N}$, where each $\boldsymbol{X}_j \in \mathbb{R}^{K\times N_j}$ can be written as $\boldsymbol{X}_j = [\boldsymbol{X}_j^1; \boldsymbol{X}_j^2; \cdots; \boldsymbol{X}_j^C]$ to satisfy $\boldsymbol{D}\boldsymbol{X}_j = \sum_{k=1}^{C} \boldsymbol{D}_k \boldsymbol{X}_j^k$. Hereafter, by convention, the concatenation of two matrices (including vectors) will be written as $[\boldsymbol{A}_1, \boldsymbol{A}_2] \doteq [\boldsymbol{A}_1\ \boldsymbol{A}_2]$ and $[\boldsymbol{A}_1; \boldsymbol{A}_2] \doteq \begin{bmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\end{bmatrix}$. The Original FDDL (O-FDDL) model is

$$\begin{aligned} &\min_{\boldsymbol{X},\boldsymbol{D}} \quad R_{\mathrm{O}}(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{X}) + \lambda_2 f(\boldsymbol{X}) + \lambda_1 \| \boldsymbol{X} \|_1 \\ &\text{s. t.} \quad \| \boldsymbol{d}_l \|_2 = 1 \quad (l = 1, \ldots, K), \end{aligned} \tag{1}$$

where

$$R_{\mathrm{O}}(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{X}) = \sum_{j=1}^{C} \left(\| \boldsymbol{Y}_j - \boldsymbol{D}\boldsymbol{X}_j \|_F^2 + \| \boldsymbol{Y}_j - \boldsymbol{D}_j \boldsymbol{X}_j^j \|_F^2 + \sum_{k\neq j} \| \boldsymbol{D}_k \boldsymbol{X}_j^k \|_F^2 \right)$$

and $f(\boldsymbol{X}) = \mathrm{Tr}(S_W(\boldsymbol{X}) - S_B(\boldsymbol{X})) + \eta \| \boldsymbol{X} \|_F^2$. Here $S_W(\boldsymbol{X})$ is the within-class scatter and $S_B(\boldsymbol{X})$ is the between-class scatter, λ_1 and λ_2 are the sparsity regularization parameter and the regularization parameter associated with f, respectively.

In (1), minimizing the term $\| \boldsymbol{Y}_j - \boldsymbol{D}_j \boldsymbol{X}_j^j \|_F^2 + \sum_{k\neq j} \| \boldsymbol{D}_k \boldsymbol{X}_j^k \|_F^2$ emphasizes the principle that the learned dictionary $\boldsymbol{D}$, which is a concatenation of class-specific sub-dictionaries $\boldsymbol{D}_j$ with $j = 1, 2, \ldots, C$, should represent $\boldsymbol{Y}_j$ discriminatively. And minimizing the term $\| \boldsymbol{Y}_j - \boldsymbol{D}\boldsymbol{X}_j \|_F^2$ accurately reflects the idea that the whole dictionary $\boldsymbol{D}$ should also represent $\boldsymbol{Y}_j$ collaboratively. Besides, the Fisher Discrimination Criterion (FDC) is applied to strength the discriminativeness of the representation coefficients $\boldsymbol{X}$. Specifically, minimizing the trace difference form of FDC, i.e., $\mathrm{Tr}(S_W(\boldsymbol{X}) - S_B(\boldsymbol{X}))$, is adopted. And adding the term $\eta \| \boldsymbol{X} \|_F^2$ can make $f(\boldsymbol{X})$ convex.

Algorithm 1. The O-FDDL Method.

Input: Training set $\boldsymbol{Y} \in \mathbb{R}^{n\times N} (n < N)$, initial over-complete dictionary $\boldsymbol{D}^{(0)} \in \mathbb{R}^{n\times K} (n < K)$, initial sparse representation matrix $\boldsymbol{X}^{(0)} \in \mathbb{R}^{K\times N}$, $\lambda_1 > 0$, $\lambda_2 > 0$, $\eta > 0$, the threshold value $\varepsilon > 0$ for solving problem (2), the number of iterations T for O-FDDL.

1: **Initialization**: t:=0, $\boldsymbol{X}$:=$\boldsymbol{X}^{(0)}$, $\boldsymbol{D}$:=$\boldsymbol{D}^{(0)}$.

2: **Repeat**

3: **Update** $\boldsymbol{X}$: Letting $\boldsymbol{D} = \boldsymbol{D}^{(t)}$ and computing $\boldsymbol{X}^{(t+1)}$ by solving the Fisher Discrimination based Sparse Representation (FDSR) problem:

$$\min_{\boldsymbol{X}} R_{\mathrm{O}}(\boldsymbol{Y}, \boldsymbol{D}^{(t)}, \boldsymbol{X}) + \lambda_2 f(\boldsymbol{X}) + \lambda_1 \| \boldsymbol{X} \|_1. \tag{2}$$

4: **Update** $\boldsymbol{D}$: Fixing $\boldsymbol{X} = \boldsymbol{X}^{(t+1)}$ and computing $\boldsymbol{D}^{(t+1)}$ by solving the Dictionary Update (DU) problem:

$$\min_{\boldsymbol{D}} R_{\mathrm{O}}(\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{X}^{(t+1)}) \quad \text{s. t.} \quad \| \boldsymbol{d}_l \|_2 = 1 \ \ (l = 1, \ldots, K). \tag{3}$$

5: t:=t + 1.