

Rank Constraints on Joint Dictionary Learning for Image Recognition

Haohao Meng^{1, a}, Yufeng Chen^{2, b}

¹School of Computer Science, Beijing Institute of Technology, Beijing, 100081, China;

²School of Computer Science, Beijing Institute of Technology, Beijing, 100081, China.

^amhh_2015@126.com, ^bcyf_2015@126.com

Keywords: Rank constraints, Joint dictionary learning, Image recognition.

Abstract. Sparse representation has been extensively applied to image recognition, while learning appropriate dictionaries for image content representation plays a critical role in it. An approach to simultaneously learn a common dictionary and multiple class-specific particular dictionaries achieves state-of-the-art performance. However, how to separate the particularity and commonality correctly is a quite important problem. Meanwhile, there exists an over-fitting phenomenon in the dictionary learned from given training samples which lie on a low dimensional subspace in that data samples can be linearly represented by dictionary. In this paper, we propose rank constraints on the joint dictionary learning (RC-JDL) algorithm to solve the above questions. Extensive experimental results on public available databases demonstrate the effectiveness of the proposed algorithm.

Introduction

The design of the dictionary that sparsifies the image samples is crucial for the better classification performance of sparse representation based model. It is well known that dictionary learning framework originally demands that the learned over-complete dictionary can faithfully represent the training samples [1, 2]. Wright et al. [3] directly design an over-complete dictionary whose base elements are original training samples of all categories for face recognition according to the assumption that test samples can be linearly combined by just those training samples from the same class. Nevertheless, when the number of training samples is huge, the computation of this algorithm will be a main bottleneck. To circumvent this problem, Yang et al. [4] propose a meta-face learning algorithm to obtain a class-specific sub-dictionary for each category. In order to enhance the discrimination power of each sub-dictionary, most prevailing supervised dictionary learning approaches in terms of the design of the sub-dictionaries can be roughly categorized into two main groups: (1) directly forcing the sub-dictionaries as incoherent as possible and (2) making the coding coefficients discriminative to implicitly push the discrimination power of sub-dictionaries. In the first category, Ramirez et al. [5] propose a novel penalty term to make the sub-dictionaries for each category independent so that some shared bases which learned sub-dictionaries may include do not have an effect on image classification. The second category dominates in existing supervised discriminative dictionary learning. In the inspiring work of Yang et al. [6], Fisher discrimination dictionary learning is proposed through the employment of Fisher criterion which enforces sparse representation coefficients to minimize scatter from the same class and maximize between-class scatter. In [7], Zhang and Li add a simple linear regression penalty term that demands that a vector whose non-zero element denotes label of the training image can be linearly represented by sparse coding coefficients. Based on [7], Jiang et al. [8] take label information of sub-dictionaries into consideration for face recognition. Images of all categories always share common features to some extent while those common features do not contribute to image recognition and have a bad effect when the number of training samples associated with each category is small. Inspired by the aforementioned works, an approach to simultaneously learn a common dictionary and multiple class-specific particular dictionaries achieves state-of-the-art performance for handwritten digit recognition and large-scale image classification in [9, 10, 11].

Rank Constraints on JDL

Brief Review of JDL.

Ning et al. [11] propose the joint dictionary learning (JDL) framework for large-scale visual recognition. Suppose that we have a group of c visually correlated categories, and let $A = [A_1, A_2, \dots, A_c]$ be a collection of original training samples, where A_i is training samples associated with class i th class, and $D = [D_0, D_1, D_2, \dots, D_c]$ be the visual dictionary learned from training samples, where D_0 is a common visual dictionary and D_i is the i th class-specific particular dictionary. Thus, JDL algorithm can be formulated as:

$$\min_{\{D_0, D_i, X_i\}_{i=1}^c} \sum_{i=1}^c \{ \|A_i - DX_i\|_F^2 + \lambda \|X_i\|_1 \} + \eta \varphi(X_1, X_2, \dots, X_c) \quad (1)$$

where X_i is the sparse coefficient matrix of A_i over the visual dictionary D , $\varphi(X_1, X_2, \dots, X_c)$ makes the coding coefficients discriminative to implicitly push the discrimination power of sub-dictionaries and both λ and η are a scalar constant.

Introduction of Rank Constraints.

Although high dimensional images which are always composed of millions of pixels increase the computational time and memory requirements of algorithms and have an adverse effect on the performance due to the so called curse of dimensionality, they will often be subjected to the region which has lower effective dimensionality instead of being uniformly distributed over the whole high dimensional space. Discriminative low rank dictionary learning has been proposed to get rid of the sparse noises. Therefore, we incorporate the low rank constrains on sub-dictionaries into the objective function (1) to find the effective low dimensional subspace:

$$\min_{\{D_0, D_i, X_i\}_{i=1}^c} \sum_{i=1}^c \{ \|A_i - DX_i\|_F^2 + \lambda \|X_i\|_1 \} + \eta \varphi(X_1, X_2, \dots, X_c) + \sum_{i=1}^c w_i \text{rank}([D_0, D_i]) \quad (2)$$

where $\text{rank}([D_0, D_i])$ denotes the rank of the matrix which consists of visual dictionary D_i and the common dictionary D_0 . However, $\text{rank}([D_0, D_i])$ is not trivial to solve in the procedure of optimization. The wise surrogate function is $\text{rank}(D_0) + \text{rank}(D_i)$.

$$\min_{\{D_0, D_i, X_i\}_{i=1}^c} \sum_{i=1}^c \{ \|A_i - DX_i\|_F^2 + \lambda \|X_i\|_1 \} + \eta \varphi(X_1, X_2, \dots, X_c) + \sum_{i=1}^c w_i (\text{rank}(D_0) + \text{rank}(D_i)) \quad (3)$$

There is no denying the fact that common features may not be extracted from all categories correctly in the joint dictionary learning framework. For example, when dictionary D_0 is a zero matrix, dictionary D_i can represent data samples associated with i th class itself. Therefore, we propose that the true shared subspace can be obtained by maximizing $\text{rank}(D_0)$ in this paper. As an alternative method, we minimize the D_i instead of maximizing $\text{rank}(D_0)$. Thus we derive the objective function :

$$\min_{\{D_0, D_i, X_i\}_{i=1}^c} \sum_{i=1}^c \{ \|A_i - DX_i\|_F^2 + \lambda \|X_i\|_1 \} + \eta \varphi(X_1, X_2, \dots, X_c) + \sum_{i=1}^c r_i (\text{rank}(D_i)) \quad (4)$$

where r_i is a constant parameter. We choose :

$$\varphi(X_1, X_2, \dots, X_c) = \sum_{i=1}^c \|A_i - D_i X_i^i - D_0 X_i^0\|_F^2 + \eta \sum_{i=0}^c \sum_{j=0, j \neq i}^c \|D_i^T D_j\|_F^2 \quad (5)$$

Optimization of RC-JDL.

At the first sight, the optimization procedure of RC-JDL problem (4) is not trivial to solve, but it is worth noting that it can be divided into two iterative sub-procedures: updating the sparse coefficients by fixing the dictionaries, and updating the dictionaries by fixing the sparse coefficients. Updating the sparse coefficients is solved in [10]. In the second sub-procedure, when the sparse coefficients are

fixed, sub-dictionary D_i is updated iteratively one by one. It is worth noting that there is differences in updating the class-specific dictionary and updating the common dictionary because the common dictionary has an effect on all categories of training data. Considering that the other dictionary is fixed, then the problem (4) can be converted to the following equivalent formulation:

$$\min_{D_i} \|M - D_i X^i\|_F^2 + \|N - D_i X_i^i\|_F^2 + r_i(\text{rank}(D_i)) + \eta \sum_{j=0, j \neq i}^C \|D_i^T D_j\|_F^2 + \eta \sum_{j=0, j \neq i}^C \|D_j^T D_i\|_F^2 \quad (6)$$

where $M = A - \sum_{j=0, j \neq i}^C D_j X^j$ and $N = A_i - D_0 X_i^0$. For simplicity, a new function $h(D_i)$ is defined as:

$$h(D_i) = \min_{D_i} \|M - D_i X^i\|_F^2 + \|N - D_i X_i^i\|_F^2 + \eta \sum_{j=0, j \neq i}^C \|D_i^T D_j\|_F^2 + \eta \sum_{j=0, j \neq i}^C \|D_j^T D_i\|_F^2 \quad (7)$$

the quadratic term h can be replaced with its first order approximation and a proximal term [30]. The problem (6) can be solved as follows:

$$D_i^{new} = \arg \min_{D_i} \frac{1}{2} \left\| D_i - D_i^{old} + \frac{W}{\eta_i} \right\|_F^2 + r_i(\text{rank}(D_i)) \quad (8)$$

where $2W = \frac{dh}{dD_i}$ and η_i is a constant parameter of the proximal term. The solution of problem (8)

can be found in [12]. We update the common dictionary in a similar way.

Experiments on AR Face Database.

The AR face database consists of over 4,000 frontal color images for 126 individuals. A subset of the data set containing 50 male subjects and 50 female subjects is chosen in our experiment. The images are cropped with dimension 60*43 and converted to gray scale. For each subject, the seven images with illumination and expression changes from Session 1 were used for training, and the other seven images with the same condition from Session 2 were used for testing. The direct comparison between competing methods which include SRC, three latest DL based classification methods (DKSVD dictionary learning, Fisher discrimination dictionary learning (FDDL) and joint dictionary learning (JDL)) and two popular classification methods (NN and SVM) is given in Table 1. Figure 1 illustrates the convergence of LC-JDL.

Table 1 The recognition rates of various methods on the AR database

Method	SRC	NN	DKSVD	SVM	FDDL	LC-JDL	JDL
Accuracy (%)	0.888	0.714	0.854	0.871	0.920	0.924	0.905

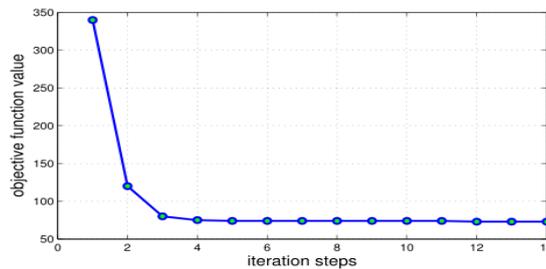


Fig. 1 The convergence of RC-JDL

Experiments on USPS hand-written digit data set.

Our second example is based on USPS hand-written digit data set which is composed of 7,291 training images and 2,007 testing images of size 16*16. Each image is represented by a 256-dimensional vector. We compare the proposed LC-JDL with several outstanding methods which include the best reconstructive DL method with linear and bilinear classifier models (denoted by REC-L and REC-BL), the best result of sparse representation for signal classification (denoted by SRSC) and the best result of FDDL and JDL. In addition, some results of problem-specific methods

(i.e., KNN and SVM with a Gaussian kernel) are also listed. Here the original image (16*16) is directly used as the feature. The comparison results are summarized in Table (2).

Table 2 Error rates of various methods on digit recognition

Method	SRSC	KNN	REC-L	SVM	FDDL	LC-JDL	REC-BL
Accuracy (%)	6.05	5.2	6.83	4.2	3.69	3.60	4.38

Summary

In this paper, we proposed a rank constraints on the joint dictionary learning (RC-JDL) approach in conjunction with sparse representation based image recognition. The RC-JDL aims to find not only the effective low dimensional subspace in which data samples lie but also the exact common subspace. The experimental results on face recognition (FR) and hand-written digit recognition obviously demonstrates the superiority of RC-JDL to many state-of-the-art dictionary learning based methods. We further plan to explore algorithm enhancements for data corrupted by missing entries and outlier and the extension of this framework for more applications.

References

- [1]. M.Aharon, M.Elad, A.Bruckstein. k-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE TSP*, 54 (11):4311–4322, 2006.
- [2]. J. Yang, K. Yu, Y. Gong, et al. Linear spatial pyramid matching using sparse coding for image classification. *CVPR*. Miami, 2009, p. 1794-1801.
- [3]. J. Wright, A. Yang, A. Ganesh, et al. Robust face recognition via sparse representation. *IEEE PAMI*. Vol. 31 (2009) No. 2, p. 210-227.
- [4]. M. Yang, L. Zhang, J. Yang, et al. Metaface learning for sparse representation based face recognition. *ICIP*. Hong Kong, 2010. p. 1601-1604.
- [5]. I. Ramirez, P. Sprechmann, G. Sapiro. Classification and clustering via dictionary learning with structured incoherence and shared features. *CVPR*. San Francisco, 2010. p. 3501-3508.
- [6]. M. Yang, D. Zhang, X. Feng. Fisher discrimination dictionary learning for sparse representation. *ICCV*. Barcelona, 2011, p. 543-550.
- [7]. Q. Zhang, B. Li. Discriminative k-svd for dictionary learning in face recognition. *CVPR*. San Francisco, 2010. p. 2691-2698.
- [8]. Z. Jiang, Z. Lin, L. S. Davis. Learning a discriminative dictionary for sparse coding via label consistent k-svd. *CVPR*. Colorado Springs, 2011, p. 1697-1704.
- [9]. S. Gao, I. W.H. Tsang, Y. Ma. Learning category specific dictionary and shared dictionary for fine-grained image categorization. *IEEE TIP*, Vol. 23 (2014) No. 2, p. 623-634.
- [10]. S. Kong, D. Wang. A dictionary learning approach for classification: separating the particularity and the commonality. *ECCV*. Florence, 2012, p. 186-199.
- [11]. N. Zhou and J. Fan. Jointly learning visually correlated dictionaries for large-scale visual recognition applications. *IEEE PAMI*, Vol. 36 (2014) No. 4, p.715-730.
- [12]. J.-F. Cai, E. J. Candès, Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, Vol. 20 (2010) No. 4, p. 1956-1982.