Neurocomputing 135 (2014) 339–347

Contents lists available at ScienceDirect

Neurocomputing

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

Low-rank decomposition and Laplacian group sparse coding for image classification

Lihe Zhang *, Chen Ma

School of Information and Communication Engineering, Dalian University of Technology, Liaoning 116023, PR China

ARTICLE INFO

Article history:
Received 29 July 2013
Received in revised form
15 November 2013
Accepted 24 December 2013
Communicated by Ran He
Available online 9 January 2014

Keywords:
Image classification
Low-rank matrix decomposition
Group sparse coding
Laplacian constraint

ABSTRACT

This paper presents a novel image classification framework (referred to as LR-LGSC) by leveraging the low-rank matrix decomposition and Laplacian group sparse coding. First, motivated by the observation that local features (such as SIFT) extracted from neighboring patches in an image usually contain correlated (or common) items and specific (or noisy) items, we construct a structured dictionary based on the low-rank and sparse components of local features. This dictionary has more powerful representation capability. Then, we investigate group generation for group sparse coding and introduce a Laplacian constraint to take into account the interrelation among groups, which can maintain low reconstruction errors while prompting similar samples to have similar codes. Finally, linear SVM classifier is used for the classification. The proposed method is tested on Caltech-101, UIUC-sports and Scene 15 dataset, and achieves competitive or better results than the state-of-the-art methods.

1. Introduction

Image classification is one of the fundamental problems in computer vision. In recent years, the bag-of-words (BoW) model [10] has been widely used to generate the feature representation of image and been applied to image classification [11,29,6], image retrieval [28], etc. The BoW quantizes local features into discrete visual words and counts their occurrence frequencies in the entire image. The resulting histogram is used as the image descriptor. However, the BoW discards the spatial organization information of local features, which severely limits the representation power. To increase the robustness against geometric deformations, an extension method called the spatial pyramid matching (SPM) [11] has been proposed. The SPM partitions an image into several finer segments in different scales, then computes the BoW histogram within each segment and concatenates all the histograms to form the image representation. This method shows promising performance and has been widely used in image classification.

The BoW framework commonly consists of local feature extraction, dictionary generation, feature coding and feature pooling. Among these steps, dictionary generation and feature coding are very important for image representation (see [7] for a review). Usually, the vector quantization (VQ) technique can code a local feature as its nearest dictionary base generated by an unsupervised clustering method, such as k-means. However, such hard assignment method may cause severe information loss [2,24] due to assigning a local feature to only one visual word, especially for the local features located among the support area of multiple visual words. To decrease the loss, soft constraint is introduced by assigning a local feature to more than one word with different weights. The weight vector indicates the contribution of each visual word to reconstruct the local feature. However, for a given local feature, it is not trivial to be determined that how many visual words should be assigned and what their weights are.

Instead of using VQ, Yang et al. [29] use sparse coding technique to determine the dictionary and the weights, and propose the method of spatial pyramid matching using sparse coding (ScSPM), which achieves the state-of-the-art performance for image classification. In sparse coding, each sample is dealt separately. The mutual dependence among samples is ignored. Hence, sparse coding might select quite different words even for similar samples to favor sparsity. Yu et al. [32] theoretically point out that under certain assumptions locality (i.e., nonzero coefficients are assigned to bases nearby to the encoded sample) is required for nonlinear coding, which is helpful for sparse codes to learn nonlinear functions in high dimensional space. To this end, they propose local coordinate coding (LCC) to encourage locally constrained sparse coding. Wang et al. [25] propose a locality-constrained linear coding (LLC) method, which can be seen as a fast implementation of LCC. They incorporate locality constraint instead of sparsity constraint into the objective function to encourage similar samples to have similar codes. While Gao et al. [4] improve the ScSPM by introducing a Laplacian constraint. This Laplacian sparse coding method (LScSPM) also preserves the consistence of sparse codes for similar samples. Thiagarajan et al. [23] propose a dictionary learning algorithm for local sparse coding, which takes into account the neighborhood relation between a dictionary atom and the training

* Corresponding author.
E-mail address: zhanglihe@dlut.edu.cn (L. Zhang).

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.12.032

vectors it represents. Interestingly, Bengio et al. [1] propose an extension of sparse coding using group coding. They leverage mixed-norm regularization to take into account the structure of bags of visual descriptors present in groups (the "group" is a set of samples). In order to obtain the sparsity in the level of image, they use the set of all local features in an image as a group (even all local features in a given category during dictionary learning). For group sparse coding, once a codeword has been selected to help represent one of the samples in a group, it is used to represent other samples in the same group as well.

In this work, we leverage group coding to capture the inter-relation between the sample space and the coding space, thereby achieving locally constrained feature coding. Although group coding makes the descriptors in a group have similar response and uniformity sparse pattern, its constraint is excessively rigid, which often brings about large reconstruction error and information loss and worsens the coding performance, especially when the samples in a group are not homogenous. To overcome this problem, we group local features by Normalized Cut clustering [22] rather than using spatial neighborhood, which can ensure the appearance consistencies of samples in a group. Moreover, inspired by the LScSPM [4], we further take into account the mutual dependence among groups by incorporating a Laplacian constraint of matrix form into the objective function of group coding, which can further promote the locality of feature coding.

On the aspect of dictionary, instead of using a common dictionary shared by all classes, the structured dictionary (i.e. collaborative representation) has been developed [34,26,30], which considers the distinctiveness of different features in the feature coding phase. The coding coefficients are more discriminative, thereby benefiting the final classification accuracy. Wright et al. [26] assume that the reconstruction error follows Gaussian or Laplacian distribution and use the reconstruction error associated with each class as the discriminative information for classification. Yang et al. [30] cast sparse coding as a sparsity-constrained robust regression problem, which seeks for the maximum likelihood estimation solution of the sparse coding problem. These methods are consistently used for face recognition. Since the samples for image classification usually have much larger scales than ones for face recognition, their practices of using training images as the dictionary atoms and sparsely encoding the holistic image features rather than local features are not suitable for our task.

In this work, we construct a structured dictionary for local features by leveraging the low-rank decomposition technique. This technique has been widely studied and successfully applied to many applications, such as image processing, web data mining. Zhang et al. [33] use it to decompose the ScSPM features of training images within each class into a low-rank matrix and a noise matrix, and then combine the both parts of all classes as the dictionary to encode test images for classification. Different from their approach, we observe that the local features extracted from adjacent patches in an image often contain correlated items and specific items. Therefore, we decompose the local features of each training image into a low-rank part and a sparse part. Based on the both parts, we learn a structured dictionary with feature-attribute information for all classes, which is more robust and discriminative for classification application than the one learned by using the raw local features because the former can more direct capture the correlative and particular attributes of an image separately.

The contributions of this paper can be summarized as follows:

1. We learn a structured dictionary for local features based on the low-rank and sparse components yielded by low-rank matrix decomposition, which can capture the correlative and particular attributes of local features.
2. We examine group sparse coding and propose Laplacian group sparse coding to take into account the dependence among different groups. Incorporated with reasonable group generation, we simultaneously ensure the locality constraint and small reconstruction error of feature coding.

The remainder of this paper is organized as follows: we review the related work in Section 2, and introduce the proposed algorithm in Section 3. Experimental results follow in Sections 4 and 5 give the conclusions.

2. Background

2.1. Low-rank matrix decomposition

The low-rank matrix recovery problem has been widely studied for the past few years. Given a matrix $\mathbf{M}$, the aim of this technique is recovering a low-rank component $\mathbf{L}$ and a sparse component $\mathbf{S}$, that is

$$\mathbf{M}=\mathbf{L}+\mathbf{S} \tag{1}$$

This decomposition problem can be solved by

$$\min_{\mathbf{L},\mathbf{S}} \quad rank(\mathbf{L})+\gamma\|\mathbf{S}\|_0 \qquad \text{s.t.} \quad \mathbf{M}=\mathbf{L}+\mathbf{S} \tag{2}$$

where the l_0-norm counts the number of nonzero entries in the sparse matrix and $\gamma>0$ is a parameter that trades off the rank term and the sparse term. However, this problem is non-convex and NP hard. Fortunately, Candes et al. [3] point that under certain conditions the problem can be transformed to

$$\min_{\mathbf{L},\mathbf{S}} \quad \|\mathbf{L}\|_*+\gamma\|\mathbf{S}\|_1 \qquad \text{s.t.} \quad \mathbf{M}=\mathbf{L}+\mathbf{S} \tag{3}$$

where the $\|\cdot\|_*$ is the nuclear norm defined as the sum of singular values. In this paper, we use inexact version of the augmented Lagrange multiplier (ALM) method proposed by Lin et al. [13] to solve the problem.

2.2. Sparse coding and group sparse coding

Based on the findings that natural images can be generally coded by structural primitives (e.g., edges and line segments) that are qualitatively similar in form to simple cell receptive fields [19,3], sparse coding selects the least possible basis from an over-complete dictionary to represent the signal of images under certain reconstruction error constraints. Intuitively, the sparsity of the coding coefficients can be measured by l_0 -norm, which counts the number of nonzero entries in a vector or matrix. Since l_0-norm regularization is an NP-hard problem, l_1-norm regularization is widely employed in sparse coding, as it is shown that l_0-norm and l_1-norm regularization are equivalent under certain conditions [13].

Let $\mathbf{Y}=[\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_k]\in\mathbf{R}^{m\times k}$ be a set of m-dimensional samples. Let $\mathbf{D}=[\boldsymbol{d}_1,\boldsymbol{d}_2,\ldots,\boldsymbol{d}_n]\in\mathbf{R}^{m\times n}$ be the dictionary with n entries. The corresponding sparse codes $\mathbf{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_k]\in\mathbf{R}^{n\times k}$ over dictionary $\mathbf{D}$ can be computed by solving the following optimization objective:

$$\min_{\mathbf{X}} \sum_{i=1}^{k} \|\boldsymbol{y}_i-\mathbf{D}\boldsymbol{x}_i\|^2+\lambda\|\boldsymbol{x}_i\|_1 = \min_{\boldsymbol{x}} \sum_{i=1}^{k} \boldsymbol{x}_i^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{D}\boldsymbol{x}_i-2\boldsymbol{y}_i^{\mathrm{T}}\mathbf{D}\boldsymbol{x}_i+\lambda\|\boldsymbol{x}_i\|_1+\boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{y} \tag{4}$$

where the parameter λ balances reconstruction quality term and sparse term. It is well known that the regularization l_1 induces sparsity and makes the problem tractable [18].