Engineering Applications of Artificial Intelligence 26 (2013) 2608–2612

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Superpixel-wise semi-supervised structural sparse coding classifier for image segmentation

Shuyuan Yang *,[1], Yuan Lv, Yu Ren, Licheng Jiao [2]

Xidian University, Taibai road, 2#, Xi'an, China

ARTICLE INFO

Article history:
Received 26 February 2013
Received in revised form
14 May 2013
Accepted 3 July 2013
Available online 20 August 2013

Keywords:
Superpixel-wise sparse coding classifier
Spatial constraint
Semi-supervised

ABSTRACT

Sparse coding based classifier (SCC) proves to lead to the state-of-the-art result in pattern recognition. Compared with traditional generative models and discriminative models, it neither casts some assumption on the distribution of data, nor learns a hyperplane to separate samples. However, SCC is characteristic of slow prediction because an l_0-norm minimization need to be solved to assign the label for each sample. In this paper, we propose a Superpixel-wise Structural Sparse Coding based Classifier (S3CC) for image segmentation. An unsupervised superpixel segmentation is first used to derive the initial labeled samples, and SCC is extended to the semi-supervised pattern where unlabeled samples are incrementally labeled and taken as the dictionary to improve the classification accuracy. Moreover, a neighborhood spatial constraint is cast on the prediction of pixel labels, to avoid the speckle-like mis-segmentation of images. Some experiments are taken on some artificial texture images, to investigate the segmentation result of our proposed S3CC. Some aspects including (1) Comparison of S3CC with SCC, (2) Comparisons of S3CC with and without spatial constraint, (3) Comparison of S3CC with semi-supervised S3CC, are tested, and the results prove the efficiency and superiority of S3CC to its counterparts.

1. Introduction

Sparse signal representation has proven to be an extremely powerful tool for acquiring, representing, and compressing high-dimensional signals. In a recent work (Wright et al., 2009), the authors proposed a Sparse Coding based Classifier (SCC) for robust face recognition, which leads to state of the art results. SCC relies on the idea that the test sample can be represented as a linear combination of the training samples. That is, given sufficient training samples of certain object class, any new sample from the same class will approximately present as the linear combination of the training samples (Wright et al., 2009), so the classification of the new sample can be reduced to a sparse coding problem where an l_0-norm needs to be minimized.

Compared with the traditional generative models and discriminative models, SCC neither casts some assumption on the distribution of data, nor learns a hyperplane to separate samples. In words, SCC is characteristic of simple principles, robustness to noise, free of model training, so working well in many pattern recognition tasks, including face recognition (Wright et al., 2009; Xiao-Tong Shuicheng, 2010), image denoising (Shutao et al., 2012), gesture recognition (Changhong et al., 2009; Boyali and Kavakli, 2012; Dong et al., 2012) and image classification (Jinjun et al., 2010; Mairal et al., 2008). However, the prediction of SCC is very slow because an l_0-norm need to be optimized to assign the label for each sample. l_0-norm minimization is a well-known NP hard optimization problem that requires an exhaustive search to locate the optimal solution. Greedy strategies (Tropp and Gilbert, 2007; Chen et al., 1989) abandon exhaustive search in favor of a series of locally optimal single-term updates and are often used to minimize l_0-norm. They can correctly pick up atoms in the case of existing sparse solution and the selection rule is simple to understand. However, they are characteristics of heavy computation and slow convergence. Some researchers indicated that l_0-norm minimization can be replaced by a substitute l_1-norm form, and many optimization algorithms prove to be able to find the sparse coefficients to predict the sample label, such as Basis Pursuit (BP) algorithm. However, most of them are iterative-based method and time-consuming.

Image segmentation is an important task in computer vision, whose goal is to partition an image into multiple sets, with all the pixels in the single set has the same labels. It can be considered as a classification problem that SCC can solve. When SCC is used for image segmentation, for the disadvantage of SCC discussed above, the computation complexity is especially high for the l_0-norm (or l_1-norm) minimization is performed pixel by pixel. In this paper, we propose a Superpixel-wise Structural Sparse Coding Classifier

* Corresponding author.
E-mail address: syyang@xidian.edu.cn (S. Yang).
[1] (*Member, IEEE*).
[2] (*Senior Member, IEEE*).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.07.002

(S3CC) for rapider image segmentation. The structural sparsity is imposed on the coding coefficients of superpixels, and a joint sparse coding strategy is used to recover the coefficients that reveal sample labels. Because SCC works in a supervised pattern and thus requires some prototype samples, in our method the initial labeled samples are derived from unsupervised superpixel segmentation. Because of extremely few labeled samples, SCC is extended to the semi-supervised pattern where the unlabeled samples are incrementally labeled and taken as the dictionary to improve the classification accuracy of SCC. Moreover, a neighborhood spatial constraint of images is cast on the prediction of sample labels, to avoid the speckle-like misclassification of pixels. Some experiments are taken on the artificial texture images to investigate the performance of our proposed S3CC, including (1) the comparison of S3CC with SCC, (2) the comparison of S3CC with and without spatial constraint, and (3) the comparison of S3CC with semi-supervised S3CC, and the results prove the superiority of S3CC to its counterparts.

The rest of this paper is organized as follows. Section 2 details the foundation of our work. The proposed S3CC is discussed in Section 3. In Section 4, some experiments are taken to compare the proposed S3CC with other related methods. The conclusions are finally summarized in Section 5.

2. Sparse coding based classifier (SCC)

SCC assumes that each pixel can be well represented as a linear combination of some training samples, and the coefficients are remarkable for the class the test sample belongs to, while for the other class, the coefficients are near to zero. Assume a training set $\mathbf{D}$ consists of $n=\sum_{m=1}^{k} n_m$ training samples of all k object classes: $\mathbf{D}=[\mathbf{D}_1,\mathbf{D}_2,...,\mathbf{D}_k]=[\mathbf{v}_{1,1},\mathbf{v}_{1,2},...,\mathbf{v}_{k,n_k}]$, where $\mathbf{v}_{m,j}$ represents the attributes of the jth pixel in the mth class and $\mathbf{D}_m$ is the training sample of the mth object class. Taking $\mathbf{D}$ as a dictionary in the sparse coding with each column normalized, one can obtain the linear representation of $\mathbf{y}$, which is a new (test) sample from the mth object class,

$$\mathbf{y}=\mathbf{D}\boldsymbol{\alpha} \tag{1}$$

where $\boldsymbol{\alpha}=[0,...,0,\alpha_{m,1},\alpha_{m,2},..,\alpha_{m,n_m},0,...,0]^T$ is a sparse coefficient vector, whose entries are zero except those associated with the mth class. The estimation of the sparse coefficients $\hat{\boldsymbol{\alpha}}$ can be obtained by solving such an l_0- norm optimization problem in (2)

$$\begin{cases} \min\|\boldsymbol{\alpha}\|_0 \\ s.t\mathbf{y}=\mathbf{D}\boldsymbol{\alpha} \end{cases} \tag{2}$$

We consider classifying the sample through the approximation error by using the coefficients corresponding to each class. After estimating the coefficient vector $\hat{\boldsymbol{\alpha}}$, one can approximate the given test sample $\hat{\mathbf{y}}$ as

$$\hat{\mathbf{y}}_i=\mathbf{D}\delta_i(\hat{\boldsymbol{\alpha}}) \quad (i=1,...,k) \tag{3}$$

by using only the coefficients with respect to the ith class. Here $\delta_i(\hat{\boldsymbol{\alpha}})$ is a new vector whose only nonzero entries are the entries in $\boldsymbol{\alpha}$ that are associated with the ith class. Then we can predict the label of $\mathbf{y}$ according to reconstruction errors between $\mathbf{y}$ and $\hat{\mathbf{y}}_i$,

$$m=\min_i r_i(\hat{\mathbf{y}})=\|\hat{\mathbf{y}}-\mathbf{D}\delta_i(\hat{\boldsymbol{\alpha}})\|_2 \tag{4}$$

This means that the best estimation of the class for test sample is the one which can present it with minimal residuals.

3. Superpixel-wise semi-supervised structural sparse coding based classifier for image segmentation

3.1. Superpixel-wise structural sparse Coding classifier (S3CC)

Different with the pixel-by-pixel classification in SCC, we perform a classification on the superpixels of images. The superpixel segmentation has been widely studied in computer vision and machine learning (Jianbo and Malik, 2000; Levinshtein et al., 2009; Moore et al., 2008; Mori et al., 2004). Using the superpixel segmentation algorithm in Moore et al. (2008), a texture imagery can be partitioned into Psuperpixels, with each superpixel consisting of $N_p(p=1,...,P)$ pixels. Let $\mathbf{Y}_p=[\mathbf{y}_{p,1},\mathbf{y}_{p,2}...\mathbf{y}_{p,N_p}]$ represent the pth superpixel consisting of N_p pixels, and denote $\mathbf{D}=[\mathbf{v}_{1,1},\mathbf{v}_{1,2},...,\mathbf{v}_{k,n_k}]$ as the dictionary composed by $n=\sum_{m=1}^{k} n_m$ training samples. In our method, we use a joint sparse representation of $\mathbf{Y}_p$ by solving such an optimization problem,

$$\begin{cases} \min\limits_{\mathbf{A}_p}\|\mathbf{A}_p\|_{\text{column},0} \\ \text{s.t.}\mathbf{Y}_p=\mathbf{D}\mathbf{A}_p \end{cases} \tag{5}$$

where $\mathbf{A}_p=[\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,...,\boldsymbol{\alpha}_{N_p}]$ is cast on a structure sparse constraint. The nonzero coefficient of each column in $\mathbf{A}_p$ appears on the same rows, and $\|\mathbf{A}_p\|_{\text{column},0}$ represents the l_0-norm of each column. That is, all the columns in $\mathbf{Y}_p$ are jointly and sparsely represented. The simultaneous OMP (SOMP) (Tropp et al., 2006), M-FOCUSS (Rao and Engan Cotter, 2004) can be used to solve (5).

Because SCC works in a supervised pattern and thus requires some prototype samples, in our method the initial labeled samples are derived from the result of the unsupervised superpixel segmentation. An unsupervised clustering is performed on the superpixel, and the sample nearest to the clustering center in each class is selected as the initial training samples in S3CC.

3.2. Semi-supervised classification with spatial constraint

Because of extremely few labeled samples, SCC is extended to the semi-supervised pattern where unlabeled samples are incrementally labeled and taken as the dictionary to improve the classification accuracy of SCC. Moreover, a neighborhood spatial constraint of images is cast on the prediction of each superpixel, to avoid the speckle-like misclassification of images. Let $\mathbf{A}_p(i,j)$ be a superpixel of interest, which is denoted by the yellow region in Fig. 1. Consider a neighbor window $NB(p)$ that contains t_p superpixel neighbors $\mathbf{A}_q(q\in NB(p))$ of the superpixel $\mathbf{A}_p(i,j)$. Fig. 1 shows five neighbors of the superpixel $\mathbf{A}_p(i,j)$.

In the image segmentation, the labels of pixels are often continuous, which can be used as a spatial constraint in the classification. In our method, the superpixel $\mathbf{A}_p(i,j)$ and its t_p superpixel neighbors are jointly and sparsely coded by solving a

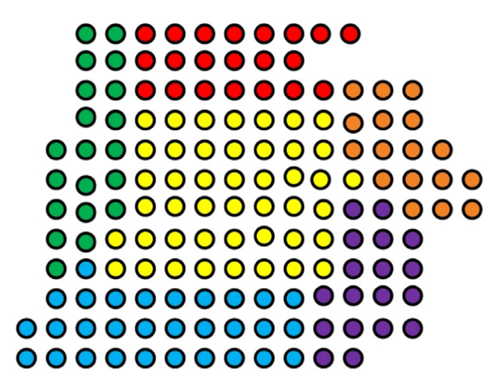

Fig. 1. The superpixel and its neighbors.