Sparse Coding for Classification via Discrimination Ensemble*

Yuhui Quan[1,3], Yong Xu[1], Yuping Sun[2,3], Yan Huang[1,3], and Hui Ji[3]
[1]School of Computer Science & Engineering, South China Univ. of Tech., Guangzhou 510006, China
[2]School of Automation Science & Engineering, South China Univ. of Tech., Guangzhou 510006, China
[3]Department of Mathematics, National University of Singapore, Singapore 117542
{csyhquan@scut.edu.cn, yxu@scut.edu.cn, ausyp@scut.edu.cn, matjh@nus.edu.sg}

Abstract

Discriminative sparse coding has emerged as a promising technique in image analysis and recognition, which couples the process of classifier training and the process of dictionary learning for improving the discriminability of sparse codes. Many existing approaches consider only a simple single linear classifier whose discriminative power is rather weak. In this paper, we proposed a discriminative sparse coding method which jointly learns a dictionary for sparse coding and an ensemble classifier for discrimination. The ensemble classifier is composed of a set of linear predictors and constructed via both subsampling on data and subspace projection on sparse codes. The advantages of the proposed method over the existing ones are multi-fold: better discriminability of sparse codes, weaker dependence on peculiarities of training data, and more expressibility of classifier for classification. These advantages are also justified in the experiments, as our method outperformed several recent methods in several recognition tasks.

1. Introduction

In recent years, as a promising technique for efficiently representing high-dimensional data, sparse coding has seen its successful usages in a variety of recognition tasks, *e.g.*, face recognition [31, 36, 3], object classification [32, 17, 3], texture classification [26, 25], and action recognition [8, 39]. Given a set of input data, sparse coding aims at expressing each input data by a linear combination of only a few elements from a set of representative patterns. These representative patterns are called *atoms*, the set of all the atoms is called *dictionary*, and the coefficients of the linear combinations are called *sparse codes*. More specifically, consider a set of input signals $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_P\} \subset \mathbb{R}^N$, sparse coding is about determining a set of atoms $\{\boldsymbol{d}_1, \boldsymbol{d}_2, \dots, \boldsymbol{d}_M\} \subset \mathbb{R}^N$, together with a set of coding vectors $\{\boldsymbol{c}_1, \dots, \boldsymbol{c}_P\} \subset \mathbb{R}^M$, so that each input vector $\boldsymbol{y}_j$ can be approximated by the linear combination $\boldsymbol{y}_j \approx \sum_{\ell=1}^{M} \boldsymbol{c}_j(\ell)\boldsymbol{d}_\ell$, where most entries of $\boldsymbol{c}_j$ are zeros or close to zeros. Let $\|\cdot\|_0$ denote the pseudo-norm that counts the number of non-zero elements. Then, the classic sparse coding problem can be formulated as the following optimization problem (e.g. [1]):

$$\min_{\boldsymbol{D},\boldsymbol{C}} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{C}\|_F^2, \quad \text{s.t. } \forall i,\ \|\boldsymbol{c}_i\|_0 \leq T, \tag{1}$$

where $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \dots, \boldsymbol{d}_M] \in \mathbb{R}^{N\times M}$ denotes the dictionary to be learned, $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_P] \in \mathbb{R}^{N\times P}$ denotes a matrix containing the input samples as column vectors, $\boldsymbol{C} = [\boldsymbol{c}_1, \dots, \boldsymbol{c}_P] \in \mathbb{R}^{M\times P}$ denotes the matrix containing the corresponding coding vectors, and the parameter T controls the sparsity degree on each coding vector. Furthermore, the normalization constraint on each atom is often imposed to avoid possible unbounded solutions, which states $\|\boldsymbol{d}_j\|_2 = 1$ for all j.

It can be seen that the dictionary learned using (1) only cares about the approximation error between the input data and the resultant succinct expression. In other words, the sparse codes obtained under the learned dictionary can be viewed as the cleaned up version of the input data. One may use such sparse codes as the features for classification. However, the additional discriminative information provided by these sparse codes over the original input signals is limited when being used in complex classification tasks, as they do not take account of the discriminability needed in classification. In recent years, there have been an abundant literature on discriminative sparse coding which is to learn a dictionary whose resultant sparse codes possess improved discriminative power; see e.g. [21, 22, 20, 33, 26, 15]. The basic idea of discriminative sparse coding for classification is to include some supervised learning processes into sparse coding. Most existing approaches for discriminative sparse

*The corresponding author Yong Xu thanks the supports by National Nature Science Foundations of China (61273255 and 61070091), Engineering & Technology Research Center of Guangdong Province for Big Data Analysis and Processing (2013-1589-1-11), Project of High Level Talents in Higher Institution of Guangdong Province (2013-2050205-47) and Guangdong Technological Innovation Project (2013KJCX0010).

1063-6919/16 $31.00 © 2016 IEEE
DOI 10.1109/CVPR.2016.629