Neural Comput & Applic
DOI 10.1007/s00521-016-2764-z

ORIGINAL ARTICLE

Sparse coding and dictionary learning with class-specific group sparsity

Yuping Sun[1] · Yuhui Quan[2] · Jia Fu[3]

Received: 21 September 2016 / Accepted: 24 November 2016

Abstract In recent years, sparse coding via dictionary learning has been widely used in many applications for exploiting sparsity patterns of data. For classification, useful sparsity patterns should have discrimination, which cannot be well achieved by standard sparse coding techniques. In this paper, we investigate structured sparse coding for obtaining discriminative class-specific group sparsity patterns in the context of classification. A structured dictionary learning approach for sparse coding is proposed by considering the $\ell_{2,0}$ norm on each class of data. An efficient numerical algorithm with global convergence is developed for solving the related challenging $\ell_{2,0}$ minimization problem. The learned dictionary is decomposed into class-specific dictionaries for the classification that is done according to the minimum reconstruction error among all the classes. For evaluation, the proposed method was applied to classifying both the synthetic data and real-world data. The experiments show the competitive performance of the proposed method in comparison with several existing discriminative sparse coding methods.

Keywords Structured sparsity · Group sparse coding · Discriminative dictionary learning · Classification

✉ Jia Fu
scutjfu@outlook.com

1 School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China

2 School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

3 School of Journalism and Communication, South China University of Technology, Guangzhou 510006, China

1 Introduction

Recent studies have shown the success of sparse modeling in analyzing high-dimensional data [14, 42, 49, 52]. The basic assumption in sparse modeling is that signals of interest can be succinctly expressed in a linear manner under some suitable system. The representative elements for expressing signals are often referred to as *atoms* and the total set of all such atoms is called a *dictionary*. The computational method for sparse modeling, which aims at finding both the dictionary and the sparse coefficients from input signals, is usually called *sparse coding*. To be more specific, given a set of signals $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p] \in \mathbb{R}^{n\times p}$, sparse coding is to determine a collection of dictionary atoms $\boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_m] \in \mathbb{R}^{n\times m}$, together with a set of sparse coefficients $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_p] \in \mathbb{R}^{m\times p}$, such that each signal $\boldsymbol{y}_i$ can be well represented by a linear combination of only a few atoms taken from $\boldsymbol{D}$:

$$\boldsymbol{y}_j \approx \sum_{\ell=1}^{m} \boldsymbol{c}_j(\ell)\boldsymbol{d}_\ell,$$

where $\boldsymbol{c}_j$ is a sparse vector with most entries being zero or close to zero.

The sparse coding is often formulated as the following minimization problem:

$$\min_{\boldsymbol{D}\in\mathcal{X},\boldsymbol{C}} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{DC}\|_F^2 + \lambda\phi(\boldsymbol{C}), \tag{1}$$

where the parameter λ determines the sparsity degree of the codes, and

$$\mathcal{X} = \{\boldsymbol{X} \in \mathbb{R}^{n\times m} : \|\boldsymbol{x}_j\|_2 = 1, 1 \le j \le m\},$$

denotes the feasible set of dictionary, which ensures that all atoms in the dictionary are appropriately normalized. The