Pattern Recognition 55 (2016) 247–260

Contents lists available at ScienceDirect

Pattern Recognition

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

Supervised dictionary learning with multiple classifier integration

Yuhui Quan, Yong Xu [*,1], Yuping Sun, Yan Huang

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

ARTICLE INFO

Article history:
Received 28 March 2015
Received in revised form
11 December 2015
Accepted 25 January 2016
Available online 23 February 2016

Keywords:
Sparse coding
Supervised dictionary learning
Multiple classifier learning
Image classification

ABSTRACT

Supervised sparse coding has become a widely-used module in existing recognition systems, which unifies classifier training and dictionary learning to enforce discrimination in sparse codes. Many existing methods suffer from the insufficient discrimination when dealing with high-complexity data due to the use of simple supervised techniques. In this paper, we integrate multiple classifier training into dictionary learning to overcome such a weakness. A minimization model is developed, in which an ensemble of classifiers for prediction and a dictionary for representation are jointly learned. The ensemble of classifiers is constructed from a set of linear classifiers, each of which is associated with a group of atoms and applied to the corresponding sparse codes. Such a construction scheme allows the dictionary and all the classifiers to be simultaneously updated during training. In addition, we provide an interesting insight into label consistency from the view of multiple classifier learning by showing its relation with the proposed method. Compared with the existing supervised sparse coding approaches, our method is able to learn a compact dictionary with better discrimination and a set of classifiers with improved robustness. The experiments in several image recognition tasks show the improvement of the proposed method over several state-of-the-art approaches.

1. Introduction

In recent years, sparse models have been widely used in a variety of applications in computer vision and pattern recognition, e.g., image analysis [1,2], image processing [3–6] and image recognition [7–15]. The philosophy of sparse modeling comes from the parsimony principle which refers to representing objects using as few variables as possible [16], and the success of sparse modeling is attributed to the fact that high-dimensional data of particular types often lie on some low-dimensional manifolds. Given a set of input data, sparse modeling aims at expressing each input data by a linear combination of a few elements taken from a set of representative patterns. The representative patterns are called atoms, and the total set of patterns is called dictionary. The coefficients of the linear combination are often referred to as sparse codes.

The dictionaries for sparse modeling are usually learned from data to maximize the efficiency of sparse approximation in terms of sparsity degree, which have shown improvement over the analytic dictionaries like wavelets in signal processing; see e.g. [3,17,18]. However, it is not optimal to use these dictionaries for classification problems where not only the sparsity but also the discriminability of sparse codes are pursued.[2] To enforce discrimination in sparse codes, the *supervised dictionary learning* methods [19–32] have been proposed to learn dictionaries in a supervised manner. The main idea of these methods is to couple the process of classifier training and the process of dictionary learning, which have exhibited impressive performance in a variety of recognition tasks. But there is still plenty of room for improvement. One possibility comes from the fact that many existing approaches (e.g. [19,20,25,32]) only employ a single simple classifier in the learning process, whose discriminative power is insufficient to handle high-complexity data. This inspired us to integrate multiple classifier learning into supervised dictionary learning.

* Corresponding author.
E-mail addresses: csyhquan@scut.edu.cn (Y. Quan), yxu@scut.edu.cn (Y. Xu), sun.yp@mail.scut.edu.cn (Y. Sun), h.y47@mail.scut.edu.cn (Y. Huang).

[1] Yong Xu would like to thank 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).

[2] The dictionaries inducing discriminative sparse codes are often referred to as discriminative dictionaries.

http://dx.doi.org/10.1016/j.patcog.2016.01.028

In this paper, we propose an effective supervised dictionary learning model which integrates multiple classifier training. Together with sparsity constraints, the objective function of the proposed model involves a simple ℓ_2 reconstructive term and a novel ensemble-based discriminative term. The discrimination term is defined by the prediction error summarized from a set of multi-class linear classifiers, each of which is associated with a group of atoms and applied to the sparse codes corresponding to the atom group. The proposed discrimination term has an interesting relation with the label consistency term used in the LC-KSVD method [32]. We then present an efficient numerical algorithm for solving the proposed model, in which the dictionary and classifiers are simultaneously updated. Used as the sparse coding module as well as the classification module, the proposed method is evaluated in several image recognition tasks, including the classification on faces, objects, scenes, actions, and dynamic textures. The experimental results have demonstrated the power of our method in discriminative sparse coding for classification.

1.1. Related work

As the goal of this paper is to develop of a dictionary learning method for sparse coding, we first give a detailed literature review on sparse dictionary learning. Then, we give a brief review on some multiple classifier learning methods which are related to our work.

1.1.1. Sparse dictionary learning

In the past, a large number of sparse models have been proposed and studied for visual recognition, whose applications cover building codebooks for local image descriptors [9,33], learning image patch representations [19,26], feature selection [34], and classification [7,27]; see [16] for a comprehensive review. This paper focuses on sparse coding and dictionary learning for classification.

The power of sparse coding for classification stems from its capability for modeling particular types of signals. There are two main successful strategies for exploiting such a capability for classification. The first one learns a class-specific dictionary for each category of signals and classifies signals by comparing the reconstruction errors or sparsity obtained under the learned dictionary. Such a classification scheme is similar in spirit to the nearest neighbor classification and the nearest subspace classification. One seminal work is the SRC method [8] that constructs the dictionaries using training samples, which has shown success in face recognition.

However, the SRC method requires a large dictionary for guaranteed performance, which is infeasible in practice due to the heavy computational burden. Such a drawback can be overcome by learning small-size dictionaries instead of simply taking signals as atoms; see e.g. [19,29,35]. But learning class-specific dictionaries separately might cause ambiguities among the learned dictionaries, i.e., signals of some class may also be well represented by the dictionaries of other classes. Several approaches have been proposed to reduce such ambiguities. Mairal et al. [19] incorporated a discriminant defined on class-specific reconstruction errors into dictionary learning to enforce the discrimination of class-specific dictionaries. Ramirez et al. [26] encouraged the independence of class-specific dictionaries by prompting the mutual incoherence among the learned dictionaries, and discarded the shared atoms which have high coherences during classification. Yang et al. [27] proposed to jointly learn class-specific dictionaries by simultaneously regarding the global and intra-class reconstruction errors and the inter-class projection energy. In [36–38], an additional global dictionary is jointly learned with class-specific dictionaries, which improves the compactness and discrimination of class-specific dictionaries. It is worth mentioning that a generalization of learning class-specific dictionaries is the so-called structured dictionary learning, which groups atoms to define structured sparsity on sparse codes. This actually allows interactions between dictionary atoms; see [39] for an example of inducing tree sparsity during dictionary learning and [14] for the concurrent image classification and annotation by grouping dictionary atoms with both class labels and image tags.

The other strategy for using sparse coding for classification is viewing dictionary atoms as discriminative features and using the corresponding sparse codes as the higher-level representations of signals for classification. The proof of this concept was first demonstrated in [7] with an analytic dictionary and a cost function built upon Fisher discriminant. Bach et al. [34] used bootstrap to improve the stability of sparse codes which is crucial to classification. For further improvement on discrimination and performance, joint cost functions that involve both a discriminative term and a classical dictionary learning formulation have been proposed. Marial et al. incorporated the softmax discriminative cost into class-specific dictionary learning [19] as well as single reconstructive dictionary learning [21]. To integrate max-margin classification into dictionary learning, the hinge loss and logistic loss are exploited in [21,23,24,40] for defining the discriminative cost. Pham et al. [20] combined the linear prediction cost with the K-SVD dictionary learning formulation for semi-supervised classification, and based on a similar model, Zhang et al. [25] developed a much more efficient algorithm. Besides the linear prediction cost, Jiang et al. [28,32] additionally considered the label consistency of subdictionaries in defining the discriminative cost, which explicitly enforces sparsity with structures under some adaptive transform and leads to impressive results in a variety of recognition tasks. As will be seen in Section 3.2, this method is very closely related to our work, and its details are presented in Section 2.2.

As the discriminative terms are constructed in the setting of supervised learning, the methods based on the second strategy are often referred to as supervised dictionary learning in the literature.[3] It is noted that optimizing a joint cost function in supervised dictionary learning requires alternating between three submodules (i.e. sparse coding, dictionary learning, and classification parameters training), which often involves a series of computationally demanding solvers and suffers from the big potential of getting stuck at local minima of the subproblems. Thus, it is preferable to develop supervised learning methods which can simultaneously update the dictionary and classification parameters. The benefits of using simultaneous update have been demonstrated in [25,28,32], where the dictionary and linear classifier are simultaneously updated using the K-SVD algorithm [3] followed by a renormalization stage. Finally, we would like to mention that supervised dictionary learning is related to neural network, as the joint process of dictionary learning and classifier construction is similar to the back propagation in network training; see [16] for details.

1.1.2. Multiple classifier learning

In the supervised setting, multiple classifier learning refers to a machine learning paradigm where a set of base classifiers is trained and combined as a strong classifier to gain extra performance [41]. It is shown in the literature that the results from the ensemble of multiple classifiers are less dependent on peculiarities

[3] In its most general definition, supervised dictionary learning also includes the dictionary learning methods using the first strategy as these methods assume class labels known.