Information Sciences 289 (2014) 25–40

Contents lists available at ScienceDirect

Information Sciences

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

How to adjust the distribution of nonzero elements in sparse representation: A granular locality-preserving approach

Yongchuan Tang *, Zejian Lee

College of Computer Science, Zhejiang University, Hangzhou 310027, PR China

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form 1 July 2014
Accepted 27 July 2014
Available online 13 August 2014

Keywords:
Sparse representation
Face recognition
Digit recognition
Letter Recognition
Granularity
Locality-preserving

ABSTRACT

In this paper, we mainly discuss the importance of distribution of nonzero elements in sparse representation. In feature space of low dimension, limited number of nonzero elements are needed to represent the target and therefore the representation is naturally sparse. Ideally, if most of nonzero elements assemble around samples of the same class as the target, the reconstruction error tends to be small and the result is more likely to be correct. Therefore, it is necessary to introduce some discriminative information into the objective function to adjust distribution of nonzero elements. We propose the Granular Locality-preserving Classification (GLC) algorithms within fine, intermediate and coarse granularity, which incorporate distance metric, class labels and clustering results of *K*-means on training data as discriminative information. Experiments conducted on several benchmark data sets show that GLC algorithms are comparable with state-of-the-art classification methods.

1. Introduction

In recent years, sparse representation has been a popular research field in artificial intelligence, pattern recognition and related areas. This trend is inspired by studies in neuroscience which find that visual image is encoded by visual cortex of human brain with a sparse representation [30,22]. Different from the common meaning, in the rest of this paper, we refer to "sparse" when we mean it is primarily populated with zeros in certain matrix, vector, etc. Namely, sparsity refers to the fraction of zero elements in a matrix or vector [28].

Due to advance of $\ell 1$-norm minimization techniques, sparse representation, namely, to represent a target sample on basis sparsely, is widely studied and it has been proven to be a useful tool for many problems of pattern recognition, such as face recognition [33], visual tracking [20], phoneme classification [23], microaneurysm detection [39], traffic sign recognition [16] and many other problems in computer vision [32]. In [33], Wright et al. proposed the Sparse Representation-based Classification (SRC), which tries to find sparse solution of linear reconstruction of the target sample via the $\ell 1$-minimization:

$$\hat{\mu} = \arg\min_{\mu} \|\mu\|_1 \quad s.t. \|X\mu - y\|_2 \leqslant \epsilon$$

where $y = [y_1, y_2, \ldots, y_m]^T$ is the query sample and $X \in R^{m \times n}$ is the training set itself. Sometimes the constraint is too strict to be feasible, so in practice the following Lasso criterion [43] is adopted:

* Corresponding author. Tel.: +86 0571 87953695; fax: +86 0571 87951250.
E-mail addresses: tyongchuan@gmail.com (Y. Tang), zejianlee@gmail.com (Z. Lee).

http://dx.doi.org/10.1016/j.ins.2014.07.035
0020-0255/

$$\hat{\mu} = \arg\min_{\mu} \lambda\|\mu\|_1 + \|X\mu - y\|_2^2$$

The objective function above is a combination of regularization term and reconstruction error term.

At last, SRC gives the judgement by examining reconstruction from which class giving the least error. Such framework shows good performance and boasts robustness to occlusion and corruption as well as insignificant choice of feature when it takes well-registered training samples as dictionary. At the same time, it still leaves a lot of space for further imagination and improvement.

Training data is not the only choice for coding basis. Based on low-rank matrix recovery, Ma et al. [18] develop a dictionary learning algorithm which yields a discriminative low-rank dictionary as the coding basis. Wang et al. [31] select codebooks which can be learned incrementally under locality-constraint as the basis. Liu et al. [16] use group sparse coding method allowing similar feature descriptors to share similar sparse representation on groups of codebooks. What is more, combining Nonnegative Matrix Factorization (NMF) [24], Liu et al. [17] choose the projections of the target spanned in each nonnegative subspace as the basis vectors. Zhu et al. [42] pick up several most contributive samples in the first round of reconstruction as the basis in the final round. Xu et al. [34] give a similar idea which select samples of several nearby classes in the first stage as candidates for the second stage. Yang et al. [37] devise an incremental dictionary learning method to find representative prototypes able to sparsely code samples from various classes. Zhang et al. [41] propose a hierarchical sparse coding algorithm to learn dictionary of basis blocklets for architecture style recognition. Based on the assumption that intraclass variations are sharable, Deng et al. [4] select training samples and intraclass variant dictionary together as coding basis. Later, they also propose a 'prototype plus variation' representation model for sparsity based face recognition [5]. Moreover, coding can be done on training samples of each class separately, which needs multiple times of minimization, as shown in [3,21,35]. In [25], Shafiee et al. discuss the role of dictionary learning and suggest that taking all training samples as basis gives the best performance but costs much more time for classification than dictionary learning methods do.

Sparse representation is characterized by the $\ell1$-norm regularization term, but the fundamental problem about the effectiveness of sparse representation leaves open because of the lack of 'rigorous mathematical justification' [32]. Zhang et al. [40] argued that, rather than sparsity from $\ell1$-norm, SRC is powerful because of its collaborative representation (the representation of training samples from all classes) but not the sparsity from $\ell1$-norm. So they propose Collaborative Representation-based Classification (CRC_RLS) [40] with $\ell2$-norm regularization term, which enables the minimization to have analytic resolution. Naseem et al. [21] also develop Linear Regression Classification (LRC) model representing an image as combination of class-specific galleries via least-squares method. Shi et al. [26] propose similar algorithm with QR factorization. In [7], Gao et al. use a Laplacian regularization term to preserve consistency of local feature. Besides regularization term, the error term also has alternative of $\ell2$-norm. In [36], Yang et al. use MLE (maximum likelihood estimation) term to estimate the fidelity of reconstruction.

Furthermore, given the coding coefficient μ, other rules instead of Nearest Subspace [15] can be taken to identify the target y. Brown [3] introduces kernel trick to the residual term, and Zhang et al. [40] proposes the regularized residual term. In addition, Yang et al. [35] chooses the class which has the minimum of $\ell1$-norm of the coefficient as the judgement, after coding the query on each class respectively, and such rule shows good efficiency.

Different from previous works mentioned above, we mainly discuss the distribution of nonzero elements in sparse representation in this paper. More in detail, we will show how the distribution of nonzero elements in the representation vector influences efficacy of algorithm, and give our solution to adjust the distribution. What is more, we propose another simple but effective rule to judge the target based on sparse coefficients.

It is argued in [26] that the assumed linear dependence among homo-class samples in [33] does not exist, and in high dimension space $\ell1$-minimization still gives dense solution. However, in space of low dimension, only small amount of independent samples are needed to represent the target, thus sparse representation is still meaningful. In sparse representation, since only small amount of elements are nonzero, distribution of these nonzero elements plays a critical role in the classification. Ideally, if most of the nonzero elements are just corresponding to training samples of the same class as the target, i.e., the target is mainly represented by the homo-class samples, reconstruction error of this correct class is smaller and therefore we are more likely to have correct classification. Such homo-class representation is the basic pursuit of sparse representation. Thus, it is necessary to incorporate some discriminative information into the objective function so as to adjust the distribution mentioned above. But what kind of information to choose and how to incorporate the information are still open challenges.

We give our solution by designing new objective functions which introduce three kinds of granular information: metric of distances between training samples and the target, the known class labels of training set, and clustering results of K-means on training data. The reason why we take metric of distances into account is because of the assumption that samples in smaller distances are prone to be of the same class. So we hope the majority of nonzero elements of the solution are just the coefficients of training samples near the test target. To put it more explicitly, the closer the base element is to the target sample, the more probably the corresponding coefficient will be nonzero, or even the larger the value of that will be; at the same time, the remoter the base element is to the target sample, the more likely the coefficient will be zero. Distinct from previous works [31,3], we do not use selective method like K-nearest-neighbour (KNN) to enforce closeness. Instead, our algorithm is able to automatically yield locality-preserving coefficients. Furthermore, since we expect homo-class representation, we should consider the labels of training data via dividing and treating them differently by class. But label information is