Accepted Manuscript

Online Learning Sensing Matrix and Sparsifying Dictionary Simultaneously for Compressive Sensing

Tao Hong, Zhihui Zhu

PII: S0165-1684(18)30190-7
DOI: 10.1016/j.sigpro.2018.05.021
Reference: SIGPRO 6829

To appear in: *Signal Processing*

Received date: 7 November 2017
Revised date: 22 April 2018
Accepted date: 22 May 2018

Please cite this article as: Tao Hong, Zhihui Zhu, Online Learning Sensing Matrix and Sparsifying Dictionary Simultaneously for Compressive Sensing, *Signal Processing* (2018), doi: 10.1016/j.sigpro.2018.05.021

Online Learning Sensing Matrix and Sparsifying Dictionary Simultaneously for Compressive Sensing

Tao Hong[a], Zhihui Zhu[b]

[a]Department of Computer Science, Technion - Israel Institute of Technology, Haifa, 32000, Israel.
[b]Department of Electrical Engineering, Colorado School of Mines, Golden, CO 80401 USA.

Abstract

This paper considers the problem of simultaneously learning the Sensing Matrix and Sparsifying Dictionary (SMSD) on a large training dataset. To address the formuated joint learning problem, we propose an online algorithm that consists of a closed-form solution for optimizing the sensing matrix with a fixed sparsifying dictionary and a stochastic method for learning the sparsifying dictionary on a large dataset when the sensing matrix is given. Benefiting from training on a large dataset, the obtained compressive sensing (CS) system by the proposed algorithm yields a much better performance in terms of signal recovery accuracy than the existing ones. The simulation results on natural images demonstrate the effectiveness of the suggested online algorithm compared with the existing methods.

Keywords: Compressive sensing, sensing matrix design, sparsifying dictionary, large dataset, online learning

1. Introduction

Sparse representation (Sparseland) has led to numerous successful applications spanning through many fields, including image processing, machine learning, pattern recognition, and compressive sensing (CS) [1] - [6]. This model assumes that a signal $\boldsymbol{x} \in \Re^N$ can be represented as a linear combination of a few columns, also known as atoms, taken from a matrix $\boldsymbol{\Psi} \in \Re^{N \times L}$ (referred to as a dictionary):

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\theta} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{\theta} \in \Re^L$ has few non-zero entries and is the representation coefficient vector of $\boldsymbol{x}$ over the dictionary $\boldsymbol{\Psi}$ and $\boldsymbol{e} \in \Re^N$ is known as the sparse representation error (SRE) which is not nil in general case. The signal $\boldsymbol{x}$ is called K-sparse in $\boldsymbol{\Psi}$ if $\|\boldsymbol{\theta}\|_0 \leq K$ where $\|\boldsymbol{\theta}\|_0$ is used to count the number of non-zeros in $\boldsymbol{\theta}$.

The choice of dictionary $\boldsymbol{\Psi}$ depends on specific applications and can be a predefined one, e.g., discrete cosine transform (DCT), wavelet transform and a multiband modulated discrete prolate spheroidal sequences (DPSS's) dictionary [7] etc. It is also beneficial and recently widely-utilized to adaptively learn a dictionary $\boldsymbol{\Psi}$, called dictionary learning, such that a set of P training signals $\{\boldsymbol{x}_k, k = 1, 2, \cdots, P\}$ is sparsely represented by optimizing a $\boldsymbol{\Psi}$. There exist many efficient algorithms to learn a dictionary [3] and the most two popular methods among them are the method of optimal directions (MOD) [4] and the K-singular value decomposition (KSVD) algorithm [5]. In particular, we prefer to use an over-complete dictionary [5], $N < L$.

CS is an emerging framework that enables to exactly recover the signal $\boldsymbol{x}$, in which it is sparse or sparsely represented by a dictionary $\boldsymbol{\Psi}$, from a number of linear measurements that is considerably lower than the size of samples required by the Shannon-Nyquist theorem [6]. Generally speaking, researchers tend to utilize a random matrix $\boldsymbol{\Phi} \in \Re^{M \times N}$ as the sensing matrix (a.k.a projection matrix) to obtain the linear measurements

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta} + \boldsymbol{\Phi}\boldsymbol{e}, \tag{2}$$

where $M \ll N$. Abundant efforts have been devoted to optimize the sensing matrix with a predefined dictionary resulting in a CS system that outperforms the standard one (random matrix) in various cases [8] - [15].

Recently, researchers realize simultaneously optimizing sensing matrix and dictionary for the CS system yields a higher signal reconstruction accuracy than the classical CS systems which only optimize sensing matrix with a fixed dictionary [14, 15]. The main idea underlying in [14, 15] is to consider the influence of SRE in learning the dictionary (see Section 3 for the formal problem). Alternating minimization methods are introduced to jointly design the sensing matrix $\boldsymbol{\Phi}$ and the dictionary $\boldsymbol{\Psi}$ in [14, 15]. Compared to [14], closed-form solutions for updating the sensing matrix and the dictionary are derived in [15] which hence obtains a better performance in terms of signal recovery accuracy. The disadvantage of the method in [15] is that it involves many singular value decompositions (SVDs) making their algorithm inefficient in practice.

Although the method for jointly optimizing the sensing matrix and the dictionary in [14, 15] works well for a small-scale training dataset (e.g., $N = 64$ and $P = 10^4$), it becomes inefficient (and even impractical) if the dimension of the dictionary is high or the size of training dataset is very large (say with more than 10^6 patches in natural images situation) or for the case involving dynamic data like video stream. It is easy to see that the methods in [14, 15] require heavy memory and computations to

Email addresses: hongtao@cs.technion.ac.il (Tao Hong), zzhu@mines.edu (Zhihui Zhu)