Expert Systems with Applications 41 (2014) 3682–3690

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Online dictionary learning algorithm with periodic updates and its application to image denoising

Ender M. Eksioglu *

Electronics and Communications Engineering Department, Istanbul Technical University, Istanbul, Turkey

ARTICLE INFO

Keywords:
Dictionary learning
Sparse representation
Online learning
Image denoising

ABSTRACT

We introduce a coefficient update procedure into existing batch and online dictionary learning algorithms. We first propose an algorithm which is a coefficient updated version of the Method of Optimal Directions (MOD) dictionary learning algorithm (DLA). The MOD algorithm with coefficient updates presents a computationally expensive dictionary learning iteration with high convergence rate. Secondly, we present a periodically coefficient updated version of the online Recursive Least Squares (RLS)-DLA, where the data is used sequentially to gradually improve the learned dictionary. The developed algorithm provides a periodical update improvement over the RLS-DLA, and we call it as the Periodically Updated RLS Estimate (PURE) algorithm for dictionary learning. The performance of the proposed DLAs in synthetic dictionary learning and image denoising settings demonstrates that the coefficient update procedure improves the dictionary learning ability.

1. Introduction

Sparse signal representation in overcomplete dictionaries has acquired considerable interest (Kim, Chen, Kim, Pan, & Park, 2011; Plumbley, Blumensath, Daudet, Gribonval, & Davies, 2010; Rubinstein, Bruckstein, & Elad, 2010). Sparse signal representation constitutes compactly expressing a signal as a linear combination from an overcomplete set of signals or atoms. The number of atoms utilized in the linear combination is much less than the signal dimensionality, hence the sparse designation. The set of all atoms forms the redundant dictionary over which sparse representations are realized. There are a plethora of methods for sparse representation of a signal over a given dictionary (Tropp & Wright, 2010). One class of algorithms includes linear programming based optimization methods (Chen, Donoho, & Saunders, 1998). Another important class of algorithms contain the greedy methods, e.g., Orthogonal Matching Pursuit (OMP) (Tropp & Gilbert, 2007), which present computationally practical solutions to the sparse representation problem.

A subject related to sparse representation is dictionary learning (Gribonval & Schnass, 2010; Rubinstein et al., 2010; Tošić & Frossard, 2011; Yaghoobi, Blumensath, & Davies, 2009), which considers the construction of the dictionary employed for sparse coding of data. Dictionary learning examines the problem of training the atoms of a dictionary suitable for the joint sparse representation of a data set. Dictionary learning algorithms (DLAs) include Maximum Likelihood (ML) methods (Olshausen & Field, 1997), Maximum a posteriori Probability (MAP)-based methods (Kreutz-Delgado et al., 2003), the K-Singular Value Decomposition (K-SVD) algorithm (Aharon, Elad, & Bruckstein, 2006), direct optimization based methods such as Rakotomamonjy (2013) and the least-squares based Method of Optimal Directions (MOD) (Engan, Aase, & Husøy, 1999; Engan, Skretting, & Husøy, 2007). Other recent approaches to the dictionary learning problem include (Sadeghi, Babaie-Zadeh, & Jutten, 2013; Smith & Elad, 2013).

In general the previously listed methods are batch algorithms, and they process the entire data set as a batch for each iteration. Recently, online DLAs have been proposed, where the algorithm allows sequential dictionary learning as the data flows in. The online algorithms include the Recursive Least Squares (RLS)-DLA (Skretting & Engan, 2010), which is derived using an approach similar to the RLS algorithm employed in adaptive filtering. The RLS approach has also been used for sparse adaptive filtering in recent studies (Babadi, Kalouptsidis, & Tarokh, 2010; Eksioglu & Tanc, 2011). Another online DLA is the Online Dictionary Learning (ODL) algorithm of Mairal, Bach, Ponce, and Sapiro (2010).

In this paper we introduce a new DLA, which is based on the least squares solution for the dictionary estimate as is the case for the MOD algorithm and the RLS-DLA. We first present a variant of the MOD algorithm where the sparse coefficients associated with the previously seen signals are recalculated at every iteration before the dictionary is updated. This variant has much higher

* Tel./fax: +90 212 2853623.
E-mail address: eksioglue@itu.edu.tr

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.11.036

computational complexity than the MOD algorithm. We regularize this computationally expensive variant by restricting the recalculation to periodic updates. The resulting algorithm which we call as the PURE algorithm is developed by augmenting the RLS-DLA algorithm with periodic updates of the sparse representations before the dictionary estimate is formed. The PURE algorithm presents performance better than the RLS-DLA, while maintaining the same asymptotic computational complexity as the RLS-DLA and MOD algorithms. Simulations show that the introduced PURE algorithm works well in the synthetic dictionary reconstruction setting and also in image denoising applications. To the best of our knowledge this work presents the first attempt to introduce a periodic coefficient update into the two-step iterative dictionary learning procedure. Dictionary learning for given data sets results in performance improvement in various applications. These applications include but are not limited to image denoising and reconstruction (Liu, Wang, Luo, Zhu, & Ye, 2012; Wang et al., 2013; Yang, Zhao, Wang, Zhang, & Jiao, 2013) and various classification problems (Jiang, Lin, & Davis, 2013). Devising new and better dictionary learning approaches naturally leads to performance improvements in the aforementioned applications.

In the coming sections, we begin first by giving a review of dictionary learning in general, and the MOD and RLS-DLA algorithms. In Section 3 we introduce the coefficient updated version of the MOD algorithm. In Section 4, we develop a new online dictionary learning algorithm by augmenting the RLS-DLA with periodic coefficient updates. Section 5 details the computational complexity of the novel algorithms when compared to the existing methods. In Section 6 we provide detailed simulations for the novel algorithms. The simulation settings include synthetic dictionary recovery and image denoising.

2. Batch and online dictionary learning algorithms

The dictionary learning problem may be defined as finding the optimally sparsifying dictionary for a given data set. The dictionary learning problem might be formulated using different optimization objectives over a sparsity regularized cost function for a given data set. Aharon et al. (2006) suggests the following expression for constructing a sparsifying dictionary.

$$\min_{\mathbf{D},\mathbf{W}} \left\{ \sum_{n=1}^{N} \|\boldsymbol{x}_n - \mathbf{D}\boldsymbol{w}_n\|_2^2 \right\} \text{ subject to } \forall n, \quad \|\boldsymbol{w}_n\|_0 \leqslant S \tag{1}$$

or equivalently

$$\min_{\mathbf{D},\mathbf{W}} \left\{ \|\mathbf{X} - \mathbf{D}\mathbf{W}\|_F^2 \right\} \text{ subject to } \forall n, \quad \|\boldsymbol{w}_n\|_0 \leqslant S \tag{2}$$

Another similar objective for dictionary learning considered in Aharon et al. (2006) is

$$\min_{\mathbf{D},\mathbf{W}} \left\{ \sum_{n=1}^{N} \|\boldsymbol{w}_n\|_0 \right\} \text{ subject to } \forall n, \quad \|\mathbf{X} - \mathbf{D}\mathbf{W}\|_F^2 \leqslant \epsilon \tag{3}$$

$\|\cdot\|_F$ is the Frobenius norm for the matrix argument, and $\|\cdot\|_0$ is the ℓ_0 pseudo-norm for a vector argument. $\mathbf{X} \in \mathbb{R}^{M\times N}$ is the data matrix, which stores all the data vectors for time $n = 1$ through N. $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$, where N is the total number of observed data vectors, and $\boldsymbol{x}_n \in \mathbb{R}^M$ is the data vector at time n. $\mathbf{D} \in \mathbb{R}^{M\times K}$ is the dictionary matrix with K atoms as columns, that is $\mathbf{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_K]$. $\boldsymbol{w}_n \in \mathbb{R}^K$ is the sparse representation vector for $\boldsymbol{x}_n$, and $\mathbf{W} \in \mathbb{R}^{K\times N}$ is the sparse representation weight matrix, $\mathbf{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N]$. S is the maximum allowed number of nonzero elements for $\boldsymbol{w}_n$. We propose the following formulation for the sparsifying dictionary learning problem.

$$\min_{\mathbf{D},\mathbf{W}} \left\{ \|\mathbf{X} - \mathbf{D}\mathbf{W}\|_F^2 + \gamma \sum_{n=1}^{N} \|\boldsymbol{w}_n\|_0 \right\} \tag{4}$$

For appropriate selection of the parameters S, ϵ and γ, we can state that all three formulations (2)–(4) treat the dictionary learning problem in a similar manner, and they all seek the optimal dictionary which results in adequately sparse representations and an acceptable representation error for a given data record $\mathbf{X}$. The main approach utilized by the DLAs in the literature for the solution of the dictionary learning optimization problem is a two-step iterative refinement procedure. In this approach at each step of the iterations either one of $\mathbf{D}$ or $\mathbf{W}$ is held constant, and the optimization is realized over the other matrix. The ith iteration for this two-step iterative refinement approach in batch mode can be summarized as follows.

(1) Find sparse $\mathbf{W}^{(i)}$ for constant $\mathbf{D}^{(i-1)}$:

$$\boldsymbol{w}_n^{(i)} = \arg\min_{\boldsymbol{w}} \|\boldsymbol{x}_n - \mathbf{D}^{(i-1)}\boldsymbol{w}\|_2^2 + \gamma\|\boldsymbol{w}\|_0, \quad \text{for } n = 1, \ldots, N \tag{5}$$

(2) Find optimal $\mathbf{D}^{(i)}$ for constant $\mathbf{W}^{(i)}$:

$$\mathbf{D}^{(i)} = \arg\min_{\mathbf{D}} \|\mathbf{X} - \mathbf{D}\mathbf{W}^{(i)}\|_F^2 \tag{6}$$

The first step above is a batch sparse representation problem. Here, the sparse representation or vector selection problem is solved for all the N data vectors separately using the same dictionary matrix $\mathbf{D}^{(i-1)}$. The sparse representation method to apply in this step can be chosen among a multitude of methods from sparse coding literature. The sparse representation methods used by different DLAs include simple gradient descent update (Olshausen & Field, 1997), FOcal Underdetermined System Solver (FOCUSS) (Kreutz-Delgado et al., 2003), OMP (Aharon et al., 2006) and the Least Angle Regression (LARS) algorithm (Mairal et al., 2010).

The second step is where the DLAs utilizing the two step approach differ from each other. The pioneering work of Olshausen and Field (1997) suggests an ML approach, where gradient descent correction is utilized for the calculation of the updated $\mathbf{D}^{(i)}$.

$$\mathbf{D}^{(i)} = \mathbf{D}^{(i-1)} - \eta \sum_{n=1}^{N} \left(\mathbf{D}^{(i-1)}\boldsymbol{w}_n^{(i)} - \boldsymbol{x}_n \right) \boldsymbol{w}_n^{(i)T} \tag{7}$$

K-SVD (Aharon et al., 2006) uses an SVD based algorithm to update $\mathbf{D}^{(i-1)}$, where the values but not the positions of the non-zero elements of $\mathbf{W}^{(i)}$ can also get updated. The method of optimized directions or the MOD algorithm (Engan et al., 1999) has also been called as the Iterative Least Squares Dictionary Learning Algorithm or ILS-DLA (Engan et al., 2007). MOD has been proposed as a least squares iterative approach for dictionary design from data. The MOD algorithm fits into the iterative relaxation based two-step approach for dictionary design as described above. The MOD algorithm calculates the exact least squares solution for (6).

$$\mathbf{D}^{(i)} = \mathbf{X}\mathbf{W}^{(i)\dagger} = \mathbf{X}\mathbf{W}^{(i)T} \left[\mathbf{W}^{(i)}\mathbf{W}^{(i)T} \right]^{-1} \tag{8}$$

Here, $(\cdot)^\dagger$ denotes the Moore–Penrose pseudo-inverse. The outline for the MOD algorithm is presented in Algorithm 1. Here, by epoch we mean a complete run over the available training set.