ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

NCDSPICE 2016

National Conference on Digital Signal Processing, Information and Communication Engineering

Govt. Polytechnic College, Kasaragod

Vol. 5, Special Issue 4, November 2016

Sparsity Based Sequential Dictionary Learning Algorithm

Jesna Augustine[1]

Student, Applied Electronics and Instrumentation, Govt. Engineering College, Kozhikode, India [1]

Abstract: The term Sparsity refered to as the number of non zero elements in sparse approximation and it can be measured by L_0 Norm. Sparse dictionary learning algorithms used to find sparse representation of the input data in the form of linear combination of basis elements. This elements are called atoms and it compose dictionary. Dictionary learning methods have been successfully used in a number of signal and image processing applications and this includes image denoising, face recognition, compression analysis. Dictionary learning algorithms consist of two stages: a sparse coding stage and a dictionary update stage. In the first stage the dictionary is kept constant and the sparsity assumption is used to produce sparse linear approximations of the observed data. In the second stage, the coefficients of the linear combination are kept constant and the dictionary is updated to minimize a certain cost function. The performance of these methods strongly depends on the dictionary update stage since most of these methods share a similar sparse coding stage. In previous dictionary learning algorithms sparsity constraint is only used in sparse coding stage but in proposed method the sparsity constraint is used both in sparse coding and dictionary update stage.

Keywords: Sparsity, Dictionary learning, SVD.

I. INTRODUCTION

Sparse approximation of a signal is the representation that signal with less number of nonzero elements. The term sparsity refers to the number of nonzero elements in the sparse approximation of the signal. Sparsity of a signal is measured by L_0 norm. There of lot of advantages working with sparse vectors. For example calculation involving multiplying a vector by matrix takes less time to compute if the vector is sparse. Also Sparse vector requires less space when being stored on a computer as only the position and value of entries need to be stored.

Sparse dictionary learning algorithms aim to find sparse representation of the input data in the form of linear combination of basis elements. Collection of these basis elements called dictionary and each basis element is known as atoms. Dictionaries are classified into undercomplete dictionaries and overcomplete dictionaries. In undercomplete dictionaries the number of atoms is less than the dimension of the signal and in overcomplete dictionaries the number of atom is greater than the dimension of signal. Overcomplete dictionaries are the typical assumption for a sparse dictionary learning algorithm problem. Dictionary learning methods have been successfully used in a number of signal and image processing applications and it includes image denoising, face recognition, compression and FMRI data analysis.
Dictionary learning problem can be formulated as follows. Let take the given input signal as Y and $Y = [y_1\ y_2\ ....\ y_N]$, where y_i denote an element of Y, $y_i \in R^n$. Consider a given dictionary $D \in R^{n\times K}$ which contains K atoms and each atom have n dimension, $d_k \in R^n$. The dictionary learning algorithm generates a representation of signal, y_i as a sparse linear combination of the atoms d_k for $k = 1,2, \ldots K, K << N$.

$$\widehat{y_i} = Dx_i \qquad (1)$$

Where $x_i \in R^K$ is a sparse representation vector of y_i such that $\|x_i\| = s << K$. Where s is the minimum number of nonzero elements contained in x_i. This problem is shown in fig.1.

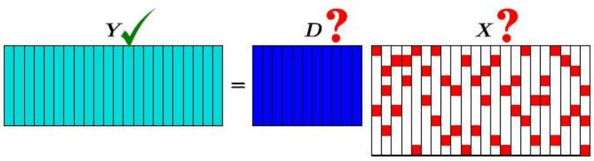

Fig 1. Dictionary learning

From figure for the given input signal we need to find the Dictionary and sparse representation of the input signal.
Dictionary learning algorithm consists of two stages: Sparse coding stage and dictionary update stage. In sparse coding stage for the given signal we try to find the sparse approximation by keeping the dictionary constant. In dictionary update stage the spare coefficients kept constant and try to find the optimum dictionary. Most of the dictionary learning algorithms is iterative between this two stages. This stage shown in fig.2.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940
IJARCCE
International Journal of Advanced Research in Computer and Communication Engineering
NCDSPICE 2016
National Conference on Digital Signal Processing, Information and Communication Engineering
Govt. Polytechnic College, Kasaragod
Vol. 5, Special Issue 4, November 2016

In sparse coding stage the dictionary is kept constant and the sparsity assumption is used to produce sparse linear approximations of the observed data. For the given input signal Y and we initialize the dictionary D with either DCT coefficients or Fourier transform coefficients or the random vectors selected from the input signal. In the proposed method the dictionary is initialized with random vectors from the input signal. In sparse coding stage we try to find the sparse approximation of input signal by using this initialized Dictionary and input signal Y. Shown in fig.3. In this stage try to minimize the error $\|Y - DX\|_F^2$.

$$\hat{x}_i = \text{argmin}_{x_i} \|y_i - Dx_i\|^2; \quad \text{subject to } \|x_i\|_0 < s$$

Different algorithms are available to compute sparse coefficients. Which includes Orthogonal Matching Pursuit(OMP), Matching Pursuit(MP), Basis Pursuit(BP) etc[5].

In dictionary update stage the coefficients of linear combination are kept constant and the dictionary is updated to minimize certain cost function. Dictionary update stage can be made sequential or parallel. In sequential approach dictionary atoms are updated sequentially and in parallel dictionary update each atom is updated in parallel. In this stage using sparse approximation of input signal which is obtained from sparse coding stage and input signal Y, obtain the optimum Dictionary D by

$$D = \text{argmin}_D \|Y - DX\|_F^2$$

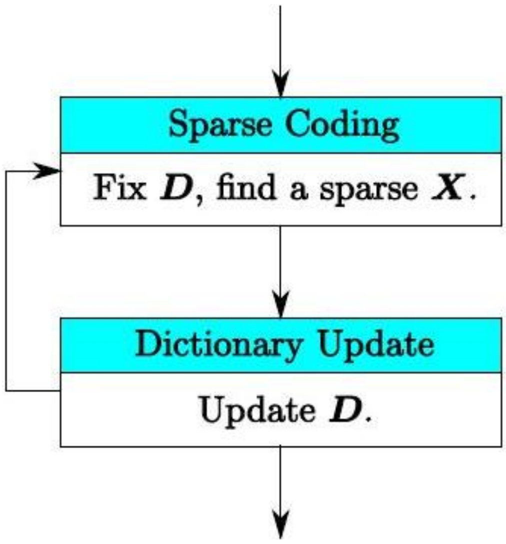

Fig. 2. Stages of Dictionary learning algorithm

The sparsity constraint used in the sparse coding stage of dictionary learning algorithm. Sparsity constraint is the pillar of any dictionary learning algorithm. While this constraint is always used in sparse coding stage, it has not been used in the dictionary update stage. In this work we are introducing the sparsity constraint also in dictionary update stage.

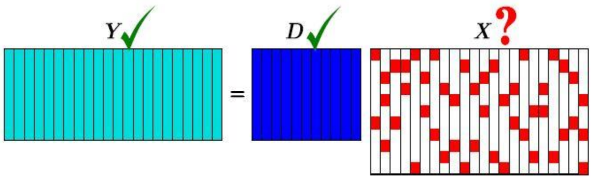

Fig. 3. Sparse Coding Stage

II. DESIGN OF DICTIONARIES: PRIOR ART

There are many existing methods available for dictionary update stage. One of them is K-SVD algorithm.

A. K-SVD Algorithm

K-SVD algorithm[1] is a sequential dictionary learning algorithm. Which means the cost function to get the optimum value of dictionary D is split into K sequential minimization. At a time only one column and the corresponding row of X is updated. And all other columns are fixed. Each column d_k of D and its corresponding row of coefficients x_k are updated based on a rank-l matrix approximation of the error for all the signals when d_k is removed. d_k and x_k are updated based on

$$\{d_k, x_k\} = \text{argmin}_{d_k, x_k^{row}} \|E_k - d_k x_k^{row}\|_F^2 \qquad (2)$$

Where E_k is the residual matrix or error matrix and it obtained by $E_k = Y - \sum_{i=1, i \neq k}^{N} d_i x_i^{row}$.

Singular value decomposition(SVD) of $E_k = U\Delta V^T$ is used to find the closest rank-1 matrix approximation of E_k. Here E_k is $n \times N$ matrix, so U is $n \times n$ matrix, V is $N \times N$ matrix and Δ forms a diagonal matrix of size $n \times N$. The d_k update is taken as first column of U and x_k^{row} update is taken as first column of V multiplied by the first element of Δ. Due to the complexity associated with computation of SVD at each stage the minimization problem in(2) can be solved by following method. The minimization problem can be rewritten as[6]

$$\begin{aligned}\text{argmin}_{d_k, x_k^{row}} \|E_k - d_k x_k^{row}\|_F^2 &= \text{tr}\{(E_k - d_k x_k^{row})(E_k - d_k x_k^{row})^T\} \\ &= \text{argmin}_{d_k, x_k^{row}} \|E_k\|_F^2 - 2d_k^T E_k x_k^{row\,T} \\ &+ \|d_k\|^2 . \|x_k^{row}\|^2\end{aligned}$$

Subject to $\|d_k\|^2 = 1$, which gives

$$d_k = \frac{E_k x_k^{row\,T}}{\|E_k x_k^{row\,T}\|_2} \quad \text{and} \quad x_k^{row} = d_k^T E_k \qquad (3)$$

Using this equations a dictionary update can be obtained by iterating (3) until convergence or by applying only one iteration of the equations instead of the computationally expensive SVD.

Direct application of SVD causes loss of sparsity on x_k^{row}. To avoid this restricts the optimization for d_k and x_k^{row}, only to the signal y_i, that use the atom d_k. Which means

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

NCDSPICE 2016

National Conference on Digital Signal Processing, Information and Communication Engineering

Govt. Polytechnic College, Kasaragod

Vol. 5, Special Issue 4, November 2016

we only update the nonzero entries in the x_k^{row}. For that an index set of $w_k = \{i|1 \le i \le N; x_k^{row}(i) \ne 0\}$ is defined. Then we define a matrix I_{w_k} as $N \times w_k$ submatrix of the $N \times N$ identity matrix obtained by retaining only those columns whose index numbers are in w_k. When multiplying $x_k^{row} \times I_{w_k}$ this shrinks the row vector x_k^{row} by discarding of the zero entries. Then $E_k^R = E_k I_{w_k}$ and taking the SVD of E_k^R will only modify the nonzero entries of x_k^{row}.

III.PROPOSED DICTIONARY UPDATE STAGE

Rather than only updating the nonzero entries of x_k^{row}, in the proposed method reupdate the sparsity of x_k^{row} in dictionary update stage. Which means here try to reduce the sparsity of x_k^{row}. To formulate this rather than minimizing optimization problem in KSVD, here update of d_k and x_k^{row} are obtained by minimization of[3]

$$\{d_k, x_k\} = \operatorname{argmin}_{d_k, x_k^{row}} \|E_k - d_k x_k^{row}\|_F^2 + \alpha \|x_k^{row}\|_1$$

Subject to $\|d_k\|_2 = 1$.

In order to reduce the sparsity of x_k^{row} here introducing some penalty parameter α[2].Which means α is a non negative penalty parameter controlling the amount of sparsity in x_k^{row}.Increasing α increases the amount of sparsity in x_k^{row}. For fixed d_k and $\|d_k\|_2 = 1$, the x_k^{row} that minimizes above equation is given by

$$x_k^{row} = \operatorname{argmin}_{x_k^{row}} \|E_k\|_F^2 - 2d_k^T E_k {x_k^{row}}^T + \|x_k^{row}\|^2 + \alpha \|x_k^{row}\|_1 \quad (4)$$

For fixed x_k^{row}, the d_k is given by

$$d_k = \operatorname{argmin}_{d_k} - 2d_k^T E_k {x_k^{row}}^T + \|d_k\|^2 . \|x_k^{row}\|^2 \quad (5)$$

Hence the solution is given by

$$x_k^{row} = \operatorname{sgn}(d_k^T E_k). \left(|d_k^T E_k| - \frac{\alpha}{2} I_{(N)}^T\right)_+ \quad (6)$$

Where $I_{(N)}$ is vector of ones of size N and

$$d_k = \frac{E_k {x_k^{row}}^T}{\left\|E_k {x_k^{row}}^T\right\|_2} \quad (7)$$

The operation in (6) is called soft-thresholding [4]. Fig.4 gives an illustration of how the soft-thresholding rule operates.

Instead of taking SVD of E_k^R to update d_k and x_k^{row} the updates of d_k and x_k^{row} are found by iterating (6) and (7) until convergence. The selection of the penalty parameter α can be obtained using a model selection criterion or cross validation. Computation cost of this iteration is O(nN).

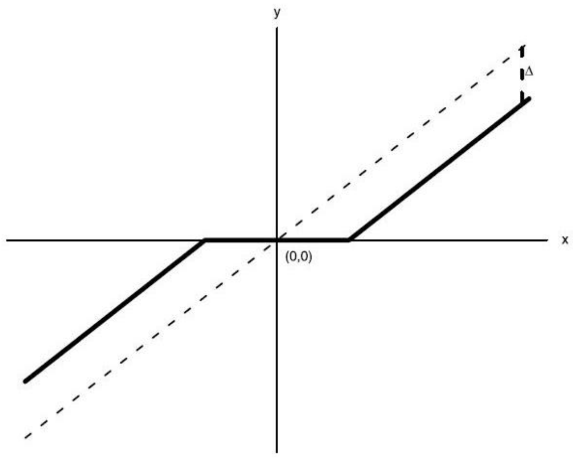

Fig. 4. An illustration of soft thresholding $\mathbf{y} = (|\mathbf{x}| - \Delta)_+ \mathbf{sgn}(\mathbf{x}), \Delta = \mathbf{1}$

IV.RESULT

Learned dictionaries are tested for the estimation of missing image data. For that 2000 patches of size 8×8 from training images form a training data Y. Using K-SVD[1] and proposed method learn dictionaries of size 64×100 from Y. Then select an input image from the set of training images. The image is divided into N non overlapping patches of size 8×8 to form image matrix $I \epsilon R^{64 \times N}$. From each image patch I_i a fraction of m random pixels are deleted, set to zero, where $m \epsilon [0.2, 0.7]$. Sparse coefficients vector is estimated using OMP for each patch with missing pixels and denoted as x_i, where i indicates the number of patch. Then reconstructed patch is obtained by

$$\widehat{I_i} = D . x_i$$

Where D is the learned dictionary. Table(1) shows the comparisons in terms of sum of squared difference(SSD) calculated from the reconstructed image and the original lena image. From table(1) we can conclude that the proposed method produce better quality estimation compared to K-SVD method.

TABLE I FILL-IN MISSING PIXELS COMPARISON IN TERMS OF SSD

Method	m= 0.3	m= 0.5
K-SVD	16.21	22.01
Proposed	15.05	20.60

V. CONCLUSION

The sparsity constraint is the pillar of any dictionary learning algorithm. Compared to previous dictionary learning algorithms in which the sparsity constraint only used in sparse coding stage, the proposed method introduce sparsity constraint both in sparse coding and dictionary update stage. Compared to state of the art methods, the proposed algorithm is computationally more efficient and generates better results.

REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein, K-SVD: An algorithm for desiging overcomplete dictionaries for sparse representation, IEEE Transactions on Signal Processing, vol. 54, pp. 4311-4322, 2006.

[2] Abd-Karim Seghouane and Muhammad Hanif, A sequential dictionary learning algorithm with enforced sparsity, ICASSP, 2015

[3] H. Shen and J. Z. Huang, Sparse principal component analysis via regularized low rank matrix approximation, Journal of Multivariate Analysis, vol. 99, pp.

[4] Hui Z, Trevor H., and Robert T.,Sparse Principal Component Analysis, Journal of Computational and Graphical Statistics, Volume 15, 2006.

[5] Roman Zeyde, Michael Elad and Matan Protter, On Single Image Scale-Up using Sparse-Representations

[6] G. H. Golub and C. f. Van Loan, Matrix Computations, Johns Hopkins, 1996.