Use Online Dictionary Learning to Get Parts-based Decomposion of Noisy Data

Daming Lu
Baidu Research
Sunnyvale, CA USA
ludaming@baidu.com

***Abstract*—Huge amount of data are generated every day. Extracting interpretable features from the data is becoming important. Meanwhile, dimension reduction and low rank approximation are also becoming important as people want to factorize big matrix into smaller ones, which are easier to handle. Sparse coding is such a technique that can factorize matrix into sparse linear combinations of basis elements. We found that through online dictionary learning, an efficient sparse coding algorithm, we could decompose large data matrix with noise into interpretable dictionary atoms. Such atoms are useful in reconstructing a denoised data matrix.**

Keywords—machine learning, sparse coding, online dictionary learning, dimension reduction

I. Introduction

Large amount of high dimensional data are generated every day, thanks to the prosperity of the Internet and big data technology. Due to the difficulty of processing high dimensional data, people intend to factorize or decompose large data matrices into smaller ones. The linear decomposition of a matrix into a few basis elements has been a hot research spot for a long time. At first, general purposed basis matrices were used to represent the large matrix, such as wavelets [1]. Later, using ad hoc matrix learned from specific input data produces better results. However, although many such decomposition methods could produce smaller matrices, these matrices are hardly interpretable, especially when the input data are noisy. Such popular methods include principal component analysis (PCA) [2], CUR matrix decomposition [3], etc. We found that online dictionary learning methods, introduced in [4], could not only reduce the matrix dimension, but the atoms in the learned dictionary are interpretable as well. We applied this technology to two different set of noisy data and found the extracted atoms very close to the ground truth.

We compared our method with *UoI NMF cluster* [5] and found our accuracies are at the same level. Meanwhile, we found that online dictionary learning runs faster as it does not require the clustering step as in *UoI NMF cluster*. The following sections are organized as below:

- We first reviewed the core part of online dictionary learning
- We then introduced the application of online dictionary learning on our datasets and compared the results with *UoI NMF cluster*
- We finally discussed potential advantages and disadvantages of both methods.
-

II. Preliminaries

A. Online Dictionary Learning

Online dictionary learning was first introduced in [4]. Assume we have a finite training dataset as $X = [x_1, \dots x_n]$ in $\mathbb{R}^{m \times n}$, we want to learn a dictionary $\mathbf{D}$ as a "good" representation of signal x. Normally the dimension m is relatively small compared to the total amount of data n. We want to have a $k \ll n$ such as we can only use a few elements (atoms) in $\mathbf{D}$ to represent signal x. Our aim is to optimize $\ell(x, \mathbf{D})$ as the ℓ_1 sparse coding problem:

$$\ell(\mathrm{x}, \mathbf{D}) \triangleq \min_{\alpha \in \mathrm{R}^k} \frac{1}{2} \|\mathrm{x} - \mathbf{D}\alpha\|_2^2 + \lambda \|\alpha\|_1$$

where λ is a regularization parameter. We followed the algorithm as mentioned in [4] with a mini-batch extension in order to prevent polluting the initialization. [5] used a clustering process to extract true bases from noise, which is also a good idea.

III. Numeric experiments

In this section, we illustrate the application of online dictionary learning on two datasets, Swimmer dataset and MNIST 2-digit dataset. The details of these two datasets were introduced in [5]. We used SPAMS library [6, 7] as our backbone. After 1,000 iterations, the performance as well as learned atoms are shown in Fig. 1 for MNIST 2-digits dataset, side by side with *UoI NMF cluster*.

We can tell from the metrics that online dictionary learning is on the same the accuracy level as *UoI NMF cluster*. Both methods learned the bases/parts pretty well. Moreover, since online dictionary learning does not have the clustering part as in *UoI NMF cluster*, it ran faster, as is shown in Table. 1.

Although *UoI NMF cluster* takes longer time, we found that it is more robust to noisy data. In both datasets, we added Gaussian noise with *sigma = 0.25*. If we tune it to be higher, *UoI NMF cluster* is more resilient whereas online dictionary learning could not extract interpretable bases. The results from Swimmer dataset are similar.

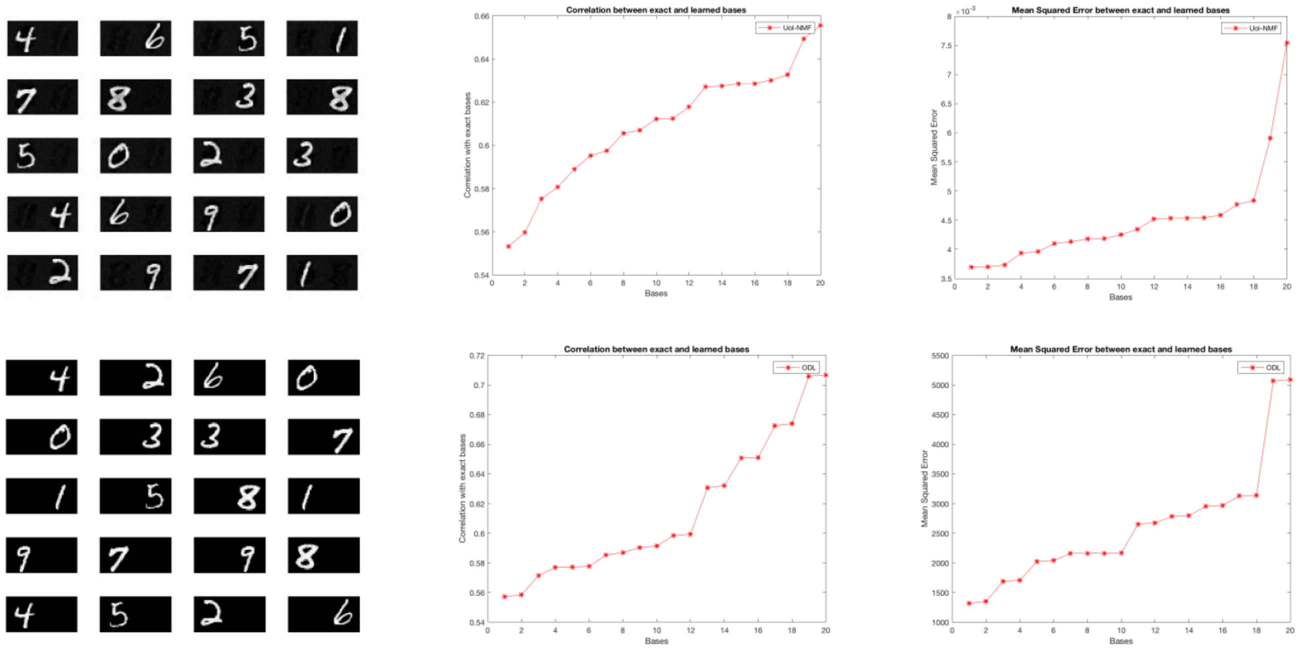

Figure 1: UoI-NMFcluster and Online Dictionary Learning for noisy MNIST two digits data. The top row includes learned bases, correlation between exact and learned bases and mean squared error between exact and learned bases, using *UoI-NMFcluster*. The bottom row includes the same metric while applying *Online Dictionary Learning*.

TABLE I. TIME AND ACCURACY

Table Head	**MNIST 2-D Dataset**	
	UoI NMF cluster	***Online Dictionary Learning***
time (sec)	109.899184	**1.747230**
reconstruction error with noisy data	**195.0384**	368.5899
reconstruction error with original data	44.6861	**16.5366**

IV. CONCLUSION

We proposed a new application of online dictionary learning to the task of parts-based decomposition of noisy data in two datasets and compared the performance with the state-of-the-art method *UoI-NMFcluster*. We found that both methods could find the actual bases in the process of dimension reduction for noisy data. *UoI-NMFcluster* takes longer time, mainly due to the DBSCAN clustering step. However, we found this step crucial for the robustness. The clustering step indeed consolidated the learned bases so that they are more robust to noise. Online dictionary learning, on the other hand, runs faster and has the potential of dealing with large amount of data, in an online or streaming way.

ACKNOWLEDGMENT

The author would like to thank Shashanka Ubaru for providing thought-provoking ideas about sparse coding, dictionary learning and nonconvex optimization.

REFERENCES

[1] Mallat, S. “A wavelet tour of signal processing”, second edition. Academic Press, New York 1999.

[2] Hervé A. “Principal component analysis”. Journal WIREs Computational Statistics Volume 2 Issue 4, July 2010 Pages 433-459

[3] Michael W. M. and Petros D. “CUR matrix decompositions for improved data analysis”. PNAS January 20, 2009 106 (3) 697-702.

[4] Julien M, Francis B., Jean P. and Guillermo S. K. Elissa, “Online Dictionary Learning for Sparse Coding” ICML 2009.

[5] Ubaru, S., Wu, K., and Bouchard, K. E. (2017). “UoI-NMFcluster: A Robust Nonnegative Matrix Factorization Algorithm for Improved Parts-Based Decomposition and Reconstruction of Noisy Data”. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 241–248.

[6] J. Mairal, F. Bach, J. Ponce and G. Sapiro. Online Learning for Matrix Factorization and Sparse Coding. Journal of Machine Learning Research, volume 11, pages 19-60. 2010.

[7] J. Mairal, F. Bach, J. Ponce and G. Sapiro. Online Dictionary Learning for Sparse Coding. International Conference on Machine Learning, Montreal, Canada, 2009.