A Multi-Scale Best Basis Sparse Learning Framework for Efficient IoT Big Data Applications

Rami Zewail[1], Ahmed Hag ElSafi[2]

[1]Smart Empower Innovation Labs Inc.,
Edmonton, Alberta, Canada
Rami.zewail@smartempower.ca

[2]Smart Empower Innovation Labs Inc.,
Edmonton, Alberta, Canada
Ahmed.hagelsafi@smartempower.ca

Abstract: *With the emergence of the Internet of Things (IoT) and cheap wireless sensors networks, IoT telemetry has received much interest over a wide range of fields including industrial applications and machine monitoring, telemedicine, and wearables. In the era of Big Data, there has been an ever-growing demand for efficient modeling and analyzing of high resolution and large-scale volumes of data. In response to these challenges, efficient signal representation has been drawing much attention lately within the research community. Concepts of sparse representation and compressed Sensing are emerging techniques that have great potentials to meet some of challenges faced by IoT and Big Data such as device power consumption, data redundancy, bandwidth, and data storage and transmission.*

In this paper we present a novel Multi-Scale Adaptive sparse learning framework for efficient signal IoT BIG DATA applications. Given a test set of the signal of interest, an adaptive dictionary learning process is proposed for sparse and redundant signal representation. First, a Joint-best basis representation is learnt in the wavelet packet domain that is optimized for the class of signals of interest. Next, dictionary learning is performed in the wavelet packet domain. The proposed framework enjoys the merits of being adaptable, efficient, and easy to deploy. The proposed framework allows for the extract of an optimal sparse representation for the signals of interest while preserving the discriminatory characteristics.

Experimental results are presented to evaluate the performance of the new framework using a wide range of signals including: faulty machinery vibration signals, ECG signals, Surface Myoelectric signals, and seismic gamma ray signals.

Keywords: Internet of Things, Big Data, sparsity, dictionary learning

1. Introduction

In the era of internet of things (IoT) and Big Data, there has been an escalating demand for efficient representation of large volumes of data. Concepts of sparse representation and sparse learning have recently drawn much attention for their potentials in efficient IoT Big Data applications. Sparse representation of a signal refers to the ability to compactly represent the signal as a linear combination of a predetermined atoms or bases (i.e. Dictionary). An efficient sparse representation model has potential to enable Big Data machine learning and artificial intelligence applications in a much lower dimension compressed domain. Hence, removing redundancy out of Big Data.

One of the most appealing applications of sparse modeling is in Compressed sensing. Compressed sensing (CS) is a technique that allow for reconstruction of signals sampled at sub-Nyquist sampling rates, given that the signal has a sparse representation in a different domain. The success of the compressed sensing application relies on the ability to project the signal into a proper sparsifying domain.

Despite the research efforts, a generic sparse representation system continues to be a challenging task. This is particularly true in IoT Big Data applications. First, the choice of the sparsifiying dictionary is crucial to the success of the sparse representation model. Moreover, to meet challenges of IoT and Big Data applications, a sparse representation model must preserve core discriminating features in the new compressed domain.

In this paper, we are concerned with the problem of efficiently representing large volumes of data by learning an adaptive, yet efficient, sparsifying dictionary. Within this context, we present a novel Multi-Scale Adaptive sparse learning framework. Given a test set of the signal of interest, an adaptive dictionary learning process is proposed for sparse and redundant signal representation. Starting with a overcomplete orthonormal wavelet packet bases, a joint best-basis representation is learnt that is optimized to the class of signals of interest. Next, K-SVD algorithm is used to perform dictionary learning in the joint best-bases wavelet packet domain. The proposed framework enjoys the merits of being adaptable, efficient, and easy to deploy. The proposed framework allows for the extract of an optimal sparse representation for the signals of interest while preserving the discriminatory characteristics.

The rest of the paper is organized as follows: Section 2 covers some of the related work in literature. Section 3 presents the details of the proposed sparse learning framework. Experiments and results are presented in section 4. Finally, conclusions are drawn in section 5.

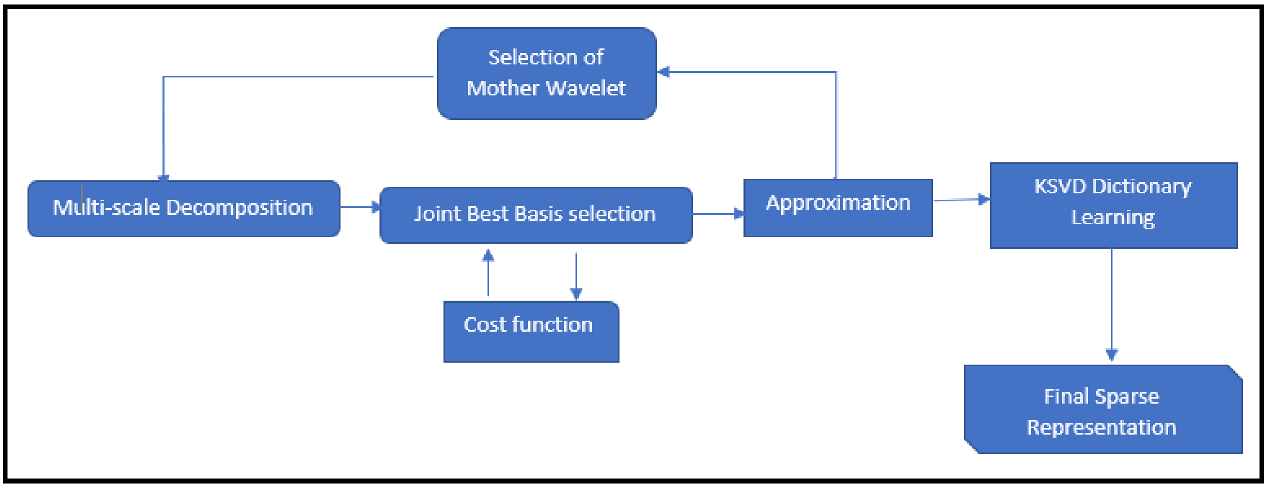

Figure 1. Block diagram for the proposed sparse learning framework

2. Related Work

The problem of efficiently representing large volumes of data has drawn much attention across different fields and applications. Over the past decade, there have been particular interest on the use of sparse representation models to solve some of the challenges of IoT Big Data such as bandwidth, storage constraints, power requirements, etc. A fundamental challenge in developing sparse signal models is the choice of a proper sparsifying dictionary. In general, there has been two general approaches for building signal sparsifying dictionaries, these are namely analytical approach and learning-based approaches.

Analytical-based approaches involves using fixed transforms such as wavelets [1], contourlets [2], curvelets [3], and others. Analytical-based approaches generally are efficient and fast but they can suffer from limited expressiveness. Learning-based approaches on the other side involves using a training set of data. Training algorithms are used to learn a set of sparsifying atom that are tailored to the training set. Examples of algorithms in this category include PCA [4], Generalized PCA [5], K-SVD [6], and others. Learning-based dictionaries are fine tuned but they suffer from complexity, and high cost.

Over past decade there has been a growing research effort to explore potentials of sparse models in some of the challenges in IoT Big Data. In [7], Mohamed Abdo-Zahhad et. al. explored the use of wavelet-based compressed sensing approach ECG telemetry applications. Riccardo Masiero,in [8], suggested using PCA as sparsifying dictionary for applications in compressed sensing in wireless sensors networks. In [9], Rajeswari C. et. al. presented a method for bearing fault diagnosis in the wavelet packet domain. In [10], Zhijin Qin et. al. explored applications of sparse representation model for wireless communications. In [11] Fausio Lucena presented a method for heart failure discrimination using matching pursuit decomposition. Yan Zhou et. al. in [12], proposed a method for speech signal compressed sensing based on K-SVD adaptive dictionary. In [13], Lingehen Zhu et. al. proposed a method to Jointly denoise and interpolate seismic data using Double-sparsity dictionary learning. In [14] , Zhilin Zhang et. al. proposed a Block Sparse Bayesian learning method for wireless telemonitoring of Noninvasive fetal ECG data. In [15] , Zhilin Zhang et. al. proposed a method for compressed sensing of

EEG for wireless telemonitoring. In [16], the author explored potential of sparse signal representation of ultrasonic signals for structural health monitoring applications. In [17], Wei Wang et. al. presented a wavelet-based analomly detection method in big data. In [24] R.Zewail et. al. presented a method for salient feature extraction using sparse contourlet-based representation. In [18], Harjeet Kaur et. al. proposed a method for detection of cardiac arrhythmia using a Wavelet-PCA based representation of ECG signals. In [25] R. Zewail and A. Hag ElSafi presented a Sparse Multi-Scale Appearance Model for pathological deformations with applications to medical imaging.

3. Methodology

This section presents the details of the proposed multiscale adaptive sparse learning framework. Figure 1 shows a block diagram of the proposed framework. Starting with a training set of signals, the proposed framework adaptively learns a sparse representation of the signal while preserving discriminating features.

Given a training set of data that represents that target signal, the signals are projected onto an overcomplete dictionary of wavelet packets. Next, we learn a joint best-basis subset of the multi-scale wavelet packet dictionary. The framework evaluates the approximation power of different mother wavelets and selects the one most suitable for the given training set. Finally, K-SVD algorithm is further used to learn a dictionary that captures the inherent characteristics of the signals in the training set.

3.1 Wavelet Packet Decomposition

We first perform wavelet packet decomposition for the training dataset. Wavelet packets represent a family of orthonormal bases in $L^2(R)$. The wavelet packet is defined as [19]:

$$\psi^{2n}(t) = \sum_k h_k \psi^n (2t-k) \qquad (1)$$

$$\psi^{2n+1}(t) = \sum_k g_k \psi^n (2t-k) \qquad (2)$$

Where: $\psi^0(t)$ is the scaling function, $\psi^1(t)$ is the wavelet function. $\{h_n\},\{g_n\}$ are the low-pass and high-pass filters associated with the wavelet transform. A wavelet packet basis is any orthonormal basis selected from the set of functions in equation 3.

$$\psi_{k,i,j} = 2^{k/2}\psi_i(2^k t - j) \qquad (3)$$

Where k=0,…,K : is the scale parameter with K as the maximum scale i = 0,.., $2^k - 1$: is the frequency index at scale k. j= 0,.., $2^{K_o - k} - 1$: is the translation index at scale k, and 2^{K_o} is the length of the input signal. The wavelet packet library can be organized in a binary tree structure, with each node representing the projection of the signal into a certain sub-space, as shown in figure 2. A p-level wavelet packet tree, divides the frequency domain into 2^p bands or sub-spaces.

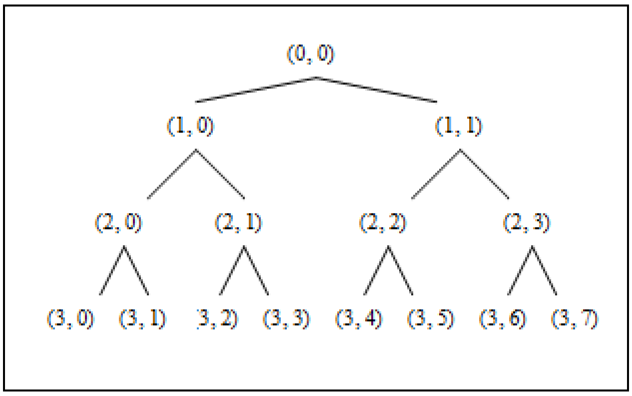

Figure 2. The tree of wavelet packet decomposition. Each doublet (k,i) represents the depth(k), and the number (i) of the nodes.

Let $\{\psi_{k,i}\}$ be the collection of all wavelet packets of translations at node (k,i). The collection of coefficients that corresponds to disjoint cover of the frequency domain will form an orthonormal basis. Thus for k levels of decomposition, there is up to 2^{2^k} bases. This flexibility offered by wavelet packets is next exploited to extract sparse, yet discriminating, representation of the signals at hand.

3.2 Finding a Joint Best Basis

Next, we search for the joint best basis from an overcomplete wavelet packet dictionary that is most suitable for the training set at hand.

For a p-level wavelet packet tree, there are 2^{2^p} possible orthonormal bases. Each basis corresponds to a specific decomposition for the frequency space. Given a library of wavelet packet basis, we search for the basis that best represent the signal in the training set.

First,we use the wavelet packet decomposition of the training data set to calculate the variance of the wavelet packet coefficients across the data set. This creates a "Variance tree" that is then used to learn the joint best basis.

Let D denotes a wavelet packet dictionary with all 2^{2^p} possible orthonormal bases $\{B_i\}$. The projection of the shape vector X onto any basis, B, is then given by:

$$C = B^T X \qquad (4)$$

Let $M(B^T X)$ denotes some information cost function for the projection of shape X onto the basis B. The best basis in the dictionary D is then the one for which the cost function, $M(B^T X)$, is minimal.

$$B_{best} = \arg\min_{B \in D}(M(B^T X)) \qquad (5)$$

Coifman and Wickehauser, [20], proposed a fast bottom-up search algorithm that finds the best orthonormal basis from a wavelet packet library according to an additive information cost function. In our model, we use the Lp norm entropy as the cost function for selecting the best basis. The Lp norm entropy for a wavelet packet node (k,i) is defined as:

$$S_L(E_j) = \sum_i \left|E_{j,i}\right|^p, 1 \le p \le 2 \qquad (6)$$

The smaller the norm entropy indicates the ability to represent localized characteristics of the signal. Hence, capturing more discriminating features among the training set. The best-basis algorithm is applied on the Variance tree to obtain a Joint-best basis set.

3.3 Selection of optimal mother wavelet

Given the training set of data, we next select the mother wavelet that is optimal for the training set.
This is done by evaluating the approximation power of the best-basis representation obtained in section 3.2 using different mother wavelets. The correlation between the original signal and the joint-best-basis -based approximated

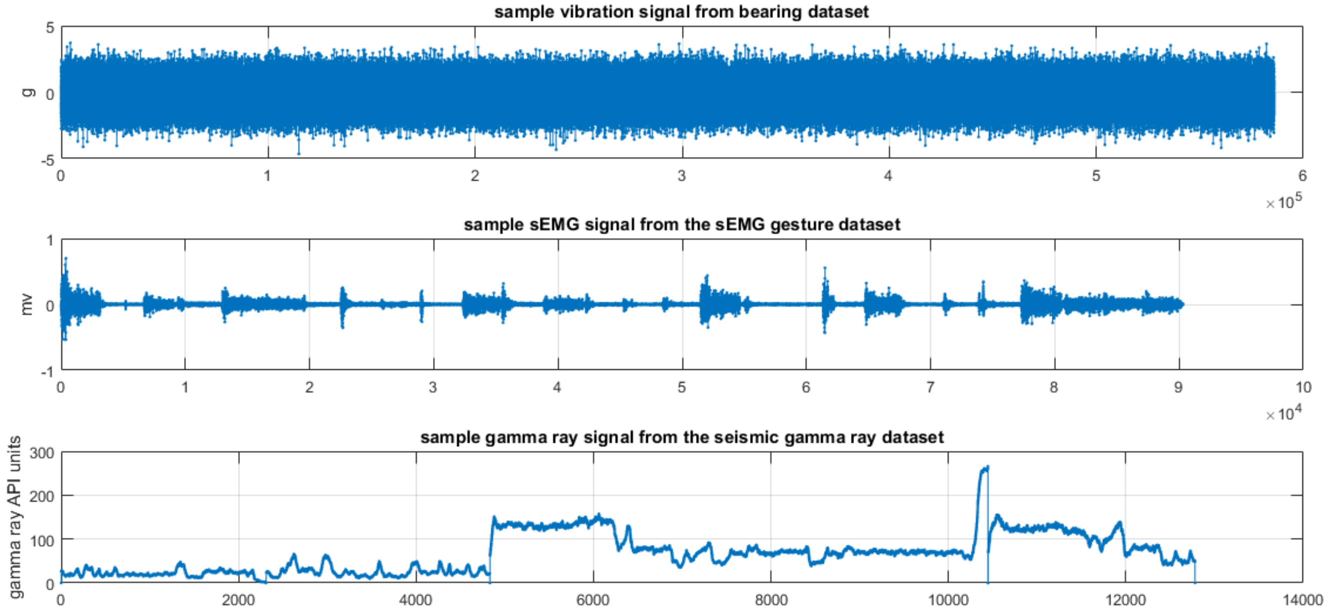

Figure 3. Example signals from the datasets used to evaluated proposed framework (a) bearing vibration signal, (b) sEMG for gesture recognition, (c) Gamma ray seismic signal

signal is used as an evaluation criterion for the mother wavelet used.

$$XCorr(X,Y) = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum (X-\overline{X})^2 (Y-\overline{Y})^2}} \quad (7)$$

3.4 K-SVD Sparse dictionary learning

The final sparse representation is obtained by using the K-SVD algorithm to learn extract the final set of sparsifying basis in the wavelet-packet domain that is capable of capturing intrinsic characteristics of the signals in the training set.

The K-SVD algorithm is a generalization for the k-means clustering. It creates a sparsifying dictionary via a singular value decomposition approach. it operates by iteratively alternating between sparse coding the input data based on the current dictionary, and updating the atoms in the dictionary to better represent the data in the training set.

The K-SVD algorithms learns an overcomplete dictionary that can sparsely represent a signal as linear combination of the dictionary atoms subject to the objective function in equation 8. Orthogonal matching pursuit (OMP) is used to calculate the coefficients.

$$\min_{D,X} \left\{ \|Y - DX\|^2 \right\}, \|x_i\| \le T_0 \quad (8)$$

4. Experimental results

This section presents the details for the experiments conducted to evaluate the sparse learning framework presented in this paper. In order to evaluate the adaptability of the proposed framework, experiments have been conducted into three different datasets of signal each with different characteristics. These are namely: machinery vibration signals, surface myoelectric physiological signals (sEMG), and gamma ray seismic signals.

4.1 Description of datasets

The first dataset used to evaluate the proposed framework is for vibration signals for a faulty bearing. The dataset is made available by the society of Machinery Failure Preventive Technology (MFPT) [21]. The original dataset is comprised of vibration signals for a normal bearing and different types of faulty bearings.

In our experiments we use a subset of the original dataset with where 50 percent of the vibration signals represent normal bearing and the other 50% are from faulty bearing with a mixture of outer and inner race faults at different loads.

The second dataset used in our evaluation process is a surface myoelectric signals (sEMG) physiological signals [22]. The dataset represents six distinct gestures. These are namely: hand open, hand close. wrist flexion. wrist extension. Supination, and pronation. Due to cheap sensors availability, recently sEMG – based gesture recognition systems have been gaining an increased attention in a variety of applications.

Finally, the third dataset used is for seismic gamma ray signals [23]. Seismic gamma ray logging is commonly used in rock and formation characterization in a borehole or drill hole. It is an extremely useful wireline method in mining, mineral exploration, and formation evaluation in oil and gas wells.

Figure 3 shows sample of the 3 different datasets used in the evaluation of the proposed framework. All three datasets used are publicly available to the research community.

4.2 Evaluation of the proposed framework

This section presents the experiments and results to evaluate the performance of the proposed sparse learning framework on the datasets described in section 4.1. The datasets are randomly split into a training and testing sets. The training set is used to perform the following learning steps:

- Learn the optimal mother wavelet suitable for the signal at hand.
- Find a joint best basis representation in the wavelet packet domain.
- Learn the final sparsifying dictionary using the K-SVD algorithm.

The testing set is then used to assess the sparsifying ability of the proposed framework, and its ability to approximate the target signal while maintaining the discriminating features in the sparse domain.

Given the training set, we next select the mother wavelet that is optimal for the training set.
Figures 4-6 show the results of the selection of the optimal mother wavelet family that is used in the joint best basis representation for the signal at hand.

Figures 4 and 5 show that the “biorthogonal 2.2” and “biorthogonal 5.5” mother wavelets were optimal for the bearing vibration signals and the physiological sEMG signals. While retaining 10% of the original coefficients, an average cross correlation of 0.7414 between the original signal and the approximated signal. An average cross correlation of 0.95 is achieved while retaining 40% of the coefficients.

Figure 5 shows that the “reverse biorthogonal 4.4” mother wavelet was optimal representing the gamma ray seismic signals. An average cross correlation of 0.95 is achieved while retraining only 10% of the coefficients.

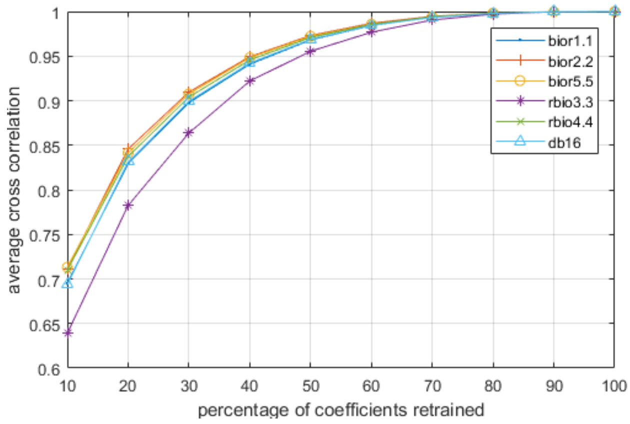

Figure 4. Selection of optimal mother wavelet for vibration bearing signals

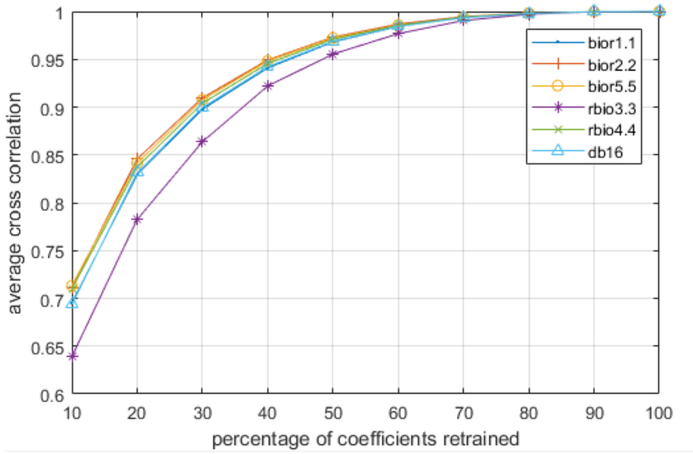

Figure 5. Selection of optimal mother wavelet for physiological sEMG signals

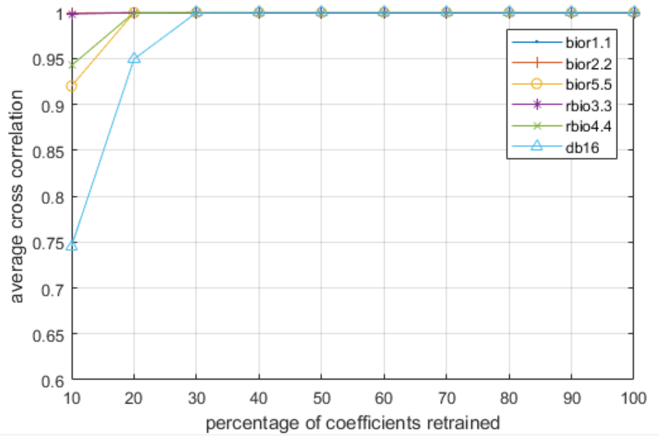

Figure 6. Selection of optimal mother wavelet for gamma ray seismic signals

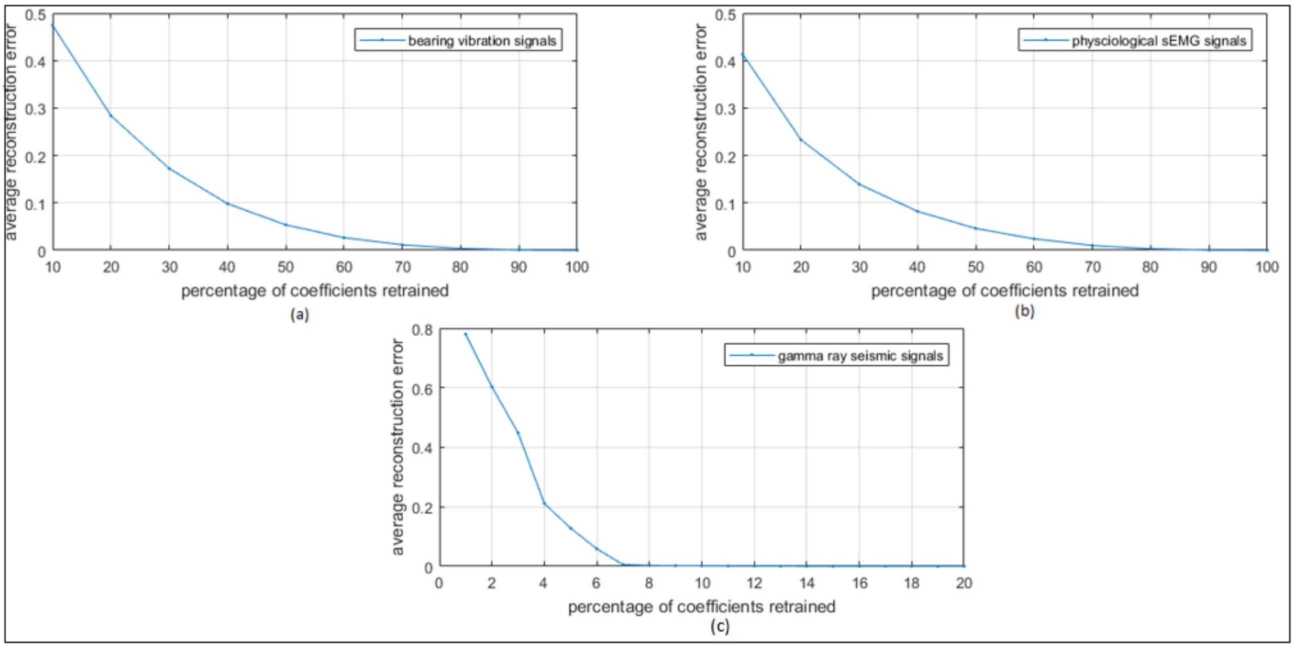

Figure 7 Approximation power for proposed sparsifying framework for different signals. (a) Bearing vibration signals, (b) Surface-EMG gesture signals, (c) gamma ray seismic signals

Figure 7 evaluates the performance of the proposed framework in terms of the approximation power vs the percentage of coefficients used. This is done for the three different signals investigated in this study. Namely, namely: machinery vibration signals, surface myoelectric physiological signals (sEMG), and gamma ray seismic signals.

For the bearing vibration signals, using the proposed framework, we were able to achieve and average reconstruction error of 0.18 using only 30% of the signal coefficients. The reconstruction error drops to 0.08 when 50% of the signal coefficients is retrained.

Figure 7.b shows that at 40% of retrained coefficients, we achieve an average reconstruction error of 0.09 for the physiological sEMG signals.

Finally, figure 7.c shows that for seismic gamma ray signals, a zero. reconstruction error was achieved using only 8% of the signals coefficients. This is very significant for applications like Measurement-while Drilling(MWD) and Logging while-drilling(LWD). As it indicates that MWD/LWD devices can potentially achieved 10 times performance improvement using the proposed sparsifying framework.

5. Summary and Conclusion

The methods proposed in this paper addressed some of the most challenging problems for IoT big Data applications. This is particularly the problem of efficiently representing large volumes of data by learning an adaptive, yet efficient, sparsifying dictionary. Within this context, we presented a novel Multi-Scale Adaptive sparse learning framework. The proposed framework enjoys the merits of being adaptable, efficient, and easy to deploy in a number of embedded platforms such as FPGAs or ARM . The proposed framework allows for the extract of an optimal sparse representation for the signals of interest while preserving the discriminatory characteristics. Experiments were conducted to evaluate the performance of the proposed framework on a wide range of signals with different intrinsic characteristics. Namely machinery vibration signals, surface myoelectric physiological signals (sEMG), and gamma ray seismic signals.

References

[1] S. Mallat, A wavelet tour of signal processing, Academic Press, 1999.

[2] M. N. Do and M. Vetterli, “The contourlet transform: an efficient directional multiresolution image representation,” IEEE Trans. Image Process., vol. 14, no. 12, pp. 2091–2106, 2005

[3] E. J. Candes and D. L. Donoho, “Curvelets – a surprisingly effective nonadaptive representation for objects with edges,” Curves and Surfaces,1999

[4] Improved EEG signal processing with wavelet based multiscale PCA algorithm.

[5] R. Vidal, Y. Ma, and S. Sastry, “Generalized principal component analysis (GPCA),” IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 12, pp. 1945–1959, 2005.

[6] M. Aharon, M. Elad, and A. M. Bruckstein, “The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation,” IEEE Trans. Signal Process. , vol. 54, no. 11, pp. 4311–4322, 2006

[7] Mohammed M. Abo-Zahhad, Aziza I. Hussein, Abdelfatah M. Mohamed, “Compression of ECG Signal Based on Compressive Sensing and the Extraction of Significant Features,” Int. J. Communications, Network and System Sciences, 2015, 8, 97-117.

[8] Masiero, Riccardo et al. “Data Acquisition through Joint Compressive Sensing and Principal Component Analysis.” GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference (2009): 1-6.

[9] C. Rajeswari, B. Sathiyabhama, S. Devendiran, and K. Manivannan, “Bearing fault diagnosis using wavelet packet transform, hybrid PSO and support vector machine*,”* Procedia Engineering, vol. 97, pp. 1772–1783, 2014.

[10] Qin, Zhijin, Fan, Jiancun; Liu, Yuanwei; Gao, Yue; Li, Geioffrey Ye, “Sparse Representation for Wireless Communications. “ In: IEEE Signal Processing Magazine, 29.12.2017.

[11] Allan Barros, Fausto Lucena, Noboru Ohnishi & Yoshinori Takeuchi, “Heart Failure Discrimination Using Matching Pursuit Decomposition ,” Proceedings of EUSIPCO, Lisbon, Portugal, 2014.

[12] Zhou, Yan and Heming. “Speech Signal Compressed Sensing Based on K- Svd Adaptive Dictionary.” (2013).

[13] L Zhu, E Liu, JH McClellan, “J*oint* Seismic *Data* Denoising and Interpolation with Double-Sparsity Dictionary Learning,” Journal of Geophysics and Engineering 14 (4), 802-810, 2017.

[14] *Zhang, Zhilin et al. “*Compressed Sensing for Energy-Efficient Wireless Telemonitoring of Noninvasive Fetal ECG Via Block Sparse Bayesian Learning.” IEEE Transactions on Biomedical Engineering 60 (2013): 300-309.

[15] Zhang, Zhilin & Jung, Tzyy-Ping & Makeig, Scott & D. Rao, Bhaskar. (2013). Compressed Sensing of EEG for Wireless Telemonitoring With Low Energy Consumption and Inexpensive Hardware. IEEE Transactions on Biomedical Engineering. 60. 221-224. 10.1109/TBME.2012.2217959.

[16] Eybpoosh, Matineh & Berges, Mario & Noh, Hae Young. (2015). Sparse representation of ultrasonic guided-waves for robust damage detection in pipelines under varying environmental and operational conditions. Structural Control and Health Monitoring. 23. 10.1002/stc.1776.

[17] Wang, W., Lu, D., Zhou, X. et al. , “Statistical wavelet-based anomaly detection in big data with compressive sensing,” J Wireless Com Network (2013) 2013: 269.

[18] Harjeet Kaur, Rajni Rjni, “ On the detection of Cardiac Arrhythmia with Principal Component Analysis,” International Journal of Wireless Personal Communiations, Kluwer Academic Publishers Vol 97 Issue 4, December, 2017

[19] M. Wickerhauser, Adapted wavelet analysis from Theory to Software. IEEE press,Piscataway, New Jersey, 1998.

[20] R. Coifman, M. Wickerhauser, “Entropy-based algorithms for best basis selection,” IEEE Transactions on Information Theory, vol. 38(2), pp. 713-718, 1992.

[21] Society of Machine Failure Preventive technology, www.mfpt.org.

[22] Chan ADC, Green GC, "Myoelectric control development toolbox", 30th Conference of the Canadian Medical & Biological Engineering Society, Toronto, Canada, M0100, 2007.

[23] “Environmental Monitoring, Other - GAMMA_RAY_LOGS_IGS_IN: Wells with Natural Gamma-Ray Logs or Physical Samples in Unconsolidated Sediments in Indiana,” www.data.gov.

[24] Rami Zewail, Ahmed Hag-ElSafi, “Appearance-based Salient Features Extraction in Medical Images Using Sparse Contourlet-based Representation”, International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.9, No.9, pp.1-10, 2017.DOI: 10.5815/ijigsp.2017.09.01

[25] Rami Zewail, Ahmed Hag-ElSafi, “MultiScale Sparse Appearance Modeling and Simulation of Pathological Deformations, “ICTACT Journal on Image and Video Processing, Vol 8, pp 1596-1605;doi: 10.21917/ijivp.2017.0225

Authors’ Profiles

Ahmed S. Hag ElSafi (Khartoum,1978) holds a BSc and a MSc of Electronics and Communications Engineering, Arab Academy for Science and Technology, 2002 and 2004 respectively.

He has 15 years of experience of research and industrial development in the areas of embedded systems and machine learning. He has published more than seventeen papers in the areas of IOT security, biometrics, machine learning, and medical image processing. He is currently a co-founder and researcher at Smart Empower Innovation Labs Inc., a Canadian-based R&D and Consulting company specialized in fields of machine learning and embedded systems.

Mr. Hag-ElSafi is a member the Smart City Alliance in Alberta, Canada and the *Association of Professional Engineers and Geoscientists of Alberta (APEGA*). He has also served as technical reviewer for Apress -Springer, NY,US

Rami Zewail received a BSc. and MSc. In Electronics & Communications Engineering, Arab Academy for Science and Technology, Egypt, in 2002 and 2004 respectively. And a PhD degree in Electrical and Computer Engineering, University of Alberta, Canada, in 2010.

He has over 15 years of academic and industrial R&D experience in areas of embedded systems, and machine learning. He has contributed to the scientific community with a patent and over 20 publications in areas of embedded computing, machine learning, and statistical modeling. Currently, he is co-founder and research at Smart Empower Innovations Labs Inc., an R&D and consulting corporation specialized in fields of embedded systems and machine learning.

Dr. Zewail is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Association of Professional Engineers & Geoscientists (APEGA), The Canadian Association for Artificial Intelligence. He also has served as a reviewer for the Journal of Electronics Imaging, Journal of Optical Engineering for the SPIE society, and Apress Media LLC-Springer.