A Data-Driven Compressive Sensing Framework for Long-Term Health Monitoring

Kai Xu, Yuhao Wang, Yixing Li, Fengbo Ren
Arizona State University

ABSTRACT

Compressive sensing (CS) is a promising technology for realizing energy-efficient wireless sensors for long-term health monitoring. In this paper, we propose a data-driven CS framework that learns signal characteristics and individual variability from patients' data to significantly enhance CS performance and noise resilience. This is accomplished by a co-training approach that optimizes both the sensing matrix and dictionary towards improved restricted isometry property (RIP) and signal sparsity, respectively. Experimental results upon ECG signals show that our framework is able to achieve better reconstruction quality with up to 80% higher compression ratio (CP) than conventional frameworks based on random sensing matrices and overcomplete bases. In addition, our framework shows great noise resilience capability, which tolerates up to 40dB higher noise energy at a CP of 9 times.

1. INTRODUCTION

The existing healthcare model based on episodic examination or short-term monitoring for disease diagnosis and treatment suffers from the overlook of individual variability and the lack of personal baseline data. Long-term or non-intermittent monitoring is the key to create the big data of individual health record for studying the variability and obtaining the personal baseline. Recent advancements in Wireless Body Area Network (WBAN) and bio-sensing techniques has enabled the emergence of miniaturized, non-invasive, cost-effective wireless sensor nodes (WSNs) that can be placed on human body for personal health monitoring [11]. Through WBAN and Internet, the monitored data can be transmitted to a near-field mobile device for on-site processing, as well as to remote servers for storage and data analysis. These technology advancements will eventually revolutionize the health related services to become more effective and economic, benefiting billions of individuals.

One of the key challenges faced by the long-term wireless health monitoring is the energy efficiency of sensing and information transfer. Due to the limited battery capacity of wireless sensor nodes (WSNs), non-intermittent sensing inevitably increases the frequency of battery recharging or replacement, making it less convenient for practical usage. In the WSNs for bio-sensing applications, the energy cost of wireless transmission is about 2 orders of magnitude greater than other components, e.g., analog-to-digital conversion (ADC) [3]. This implies that reducing the data size for information transfer is the key to improve the energy efficiency of WSNs.

Compressive sensing (CS) [2] offers a universal and simple data encoding scheme that can compress a variety of physiological signals, providing a promising solution to the above mentioned problem. However, most existing CS frameworks employ random Gaussian or Bernoulli sensing matrices that are generated independently, thereby they fail to leverage any special geometric structure embedded in the signals of interest. This limits the rank of the matrices required for achieving a bounced mutual coherence, leading to limited sensing performance. On the other hand, conventional CS frameworks [12, 6, 1] that adopt pre-determined basis for reconstruction underestimates the intricacies of philological signals and overlooks the criticality of individual variability to signal fidelity, resulting in very limited reconstruction performance [13]. Our previous study [10] has shown that overcomplete dictionaries can better approximate the underlying statistical density of input data, therefore can significantly improve the sparsity of physiological signals as well as reconstruction performance.

There have been some recent works on exploiting data structures for compressive sensing. Some aim to optimize the sensing matrix to have least coherence with the...
dictionary \([4,2]\). Yet, these methods do not guarantee distance preservation for the worst case element in the dataset. Other methods, such as NuMax \([5]\), aim to preserve pairwise distance between sample data. However, since the NuMax formulation minimizes the transformation distortion against original signal rather than its sparse coefficient, it is not compatible with overcomplete dictionaries. As a result, these methods are not suitable for the CS of physiological signals.

To overcome the limitations of existing methods, we propose a data-driven CS framework that co-trains both the sensing matrix and the dictionary to exploit the intrinsic data structure of physiological signals to enhance CS performance in this paper. We first exploits online dictionary learning (ODL) \([8]\) to train a personalized basis that further improves signal sparsity by capturing the characteristics and individual variability of physiological signals. The improved sparsity will result in better reconstruction quality and compression ratio (CP) tradeoff. Subsequently, we formulate a rank and distortion minimization problem to construct a low-rank sensing matrix that preserves the length of the worst case sparse coefficient on the trained (overcomplete) basis to further enhance the noise resilience of the sensing process.

The data-driven nature of the proposed CS framework is very appealing, because it fills the gap between the massive medical data and how to utilize them to improve the quality of monitoring. The key insight from this study is that sensor energy efficiency can be improved by learning upon data to inform compressive sensing and reconstruction through cost-effective computations on the servers, rather than doing costly circuit design. More importantly, the proposed framework has the potential to be consistently improved as more and more data is collected. Experimental results upon ECG signals show that our framework is able to achieve better reconstruction quality with up to 80% higher compression ratio (CP) than conventional frameworks based on random sensing matrices and overcomplete bases. In addition, our framework shows great noise resilience capability, which tolerates up to 40dB higher noise energy at a CP of 9 times.

3. COMPRESSION SENSING ARCHITECTURE

3.1 Architecture Overview

The architecture of the proposed framework is shown in figure 1. It is composed of three functional modules, i.e., dictionary learning (DL), optimal projection (OP), and CS signal reconstruction, performed on a server node, a sensor node, and a mobile node, respectively.

As ECG signals vary greatly among different patients, a generic basis that works for all patients can perform poorly. The dictionary learning module trains personalized basis that captures individual ECG signal features that are essential to CS recovery, which guarantees a higher sparsity than a pre-determined basis for all patients. Signal with higher sparsity requires fewer measurements for reconstruction, which further leads to power saving on sensor nodes. Before DL is performed, the original ECG signal is first preprocessed to remove baseline wandering and high-frequency interference, so that the derived dictionary is free from noise. As the signal reconstruction module finds the least number of linear combination of atoms in the dictionary, noise do not match any basis will not be reconstructed. To search for an optimum setup, we first sweep parameters used in dictionary learning. Here we employed ODL to learning a dictionary. The most notable advantage of ODL is it does not rely on the matrix factorization upon the entire training data. As a result, the time cost is much less compared to the non-online versions when handling large training datasets. As dictionary update does not depend on the previous samples, the framework also eliminates the demand of large storage space for prior inputs.

On the other hand, the sensing matrix has direct impact on the power consumption of sensor nodes. \(\Phi\) com-
presses signal from original M-dimensional space to N-dimension with N samples transmitted wirelessly to mobile node. Therefore, the rank of \( \Phi \) greatly affects the transmission power. As the OP module minimizes the rank of sensing matrix while satisfying RIP, the data transmission power of sensor node can be optimized. As the sensing matrix is optimized to guarantee RIP, we can use less measurement to achieve the same reconstruction performance as random projection even under noise environment. As the sensing matrix requires only one-time offline training, it only needs a small block of SRAM to store it, eliminating the necessity for designing extra circuit for generating pseudo random numbers, which can be power consuming.

Basic pursuit (BP) algorithm is used here for reconstruction. As the framework is compatible with other reconstruction algorithms, more computation-efficient algorithms, e.g., fast iterative shrinkage-thresholding algorithm (FISTA) can be implemented to realize real-time reconstruction.

Compared to some recent CS framework \[12, 6, 1\], the proposed one does not need to proceed the peak detection and length normalization process, which runs on the sensor nodes. So the proposed framework significantly reduced the circuit and computational complexity of sensor nodes.

### 3.2 OPTIMAL PROJECTION

In prior art, they seek to find a sensing matrix which has low coherence with the dictionary \[4\]. When the measurements are contaminated with noise or have been corrupted by some error such as quantization, it will be useful to consider somewhat stronger conditions. Candes and Tao introduced the following isometry condition on matrices \( A \) and established its importance role in CS. And if sensing matrix \( \Phi \) satisfies the RIP, then this is sufficient for a variety of algorithms to be able to successfully recover a sparse signal from noisy measurements. Here, we directly optimize RIP to allow large noise margin, which tolerates the noise from environment and quantization. RIP can be defined as:

\[
(1 - \delta)|\| \theta \|_2 \leq |\| \Phi \Psi \theta \|_2 \leq (1 + \delta)|\| \theta \|_2
\]  

(5)

where \( \theta \) is the sparse coefficients vector under dictionary \( \Psi \), \( \delta \) is the isometry constant. Inequtation (5) is equivalent to

\[
|\| \Phi \Psi \theta \|_2 - |\| \theta \|_2 | \leq \delta
\]  

(6)

As input samples are segmented into various columns, the corresponding sparse coefficients composes a matrix. Hence the optimization problem is to guarantee every column in \( \theta \) to satisfy the inequation (6) which can be defined as

\[
|\| \theta_i^T (\Psi^T \Phi^T \Phi \Psi - I) \theta_i | \leq \delta, \quad i = 1 \ldots L
\]  

(7)

where \( L \) is number of columns of sparse coefficients matrix. Assume \( A = \Phi \Psi, Y = A^T A \), equation (7) can be represented as

\[
|\| \theta_i^T (Y - I) \theta_i | \leq \delta, \quad i = 1 \ldots L
\]  

(8)

As the dimension of sensing matrix is proportional to the data size for transmission, we also optimize its rank. It means we want to find a sensing matrix to preserve the distance of every element in the dataset with least rows. While rank optimization problem is not convex, here we employ nuclear norm to relax the problem.

\[
\min_{Y} \{ |\| \theta_i^T (Y - I) \theta_i |, \text{rank}(Y) \}, \quad i = 1 \ldots L
\]

s.t. \( Y \succ 0 \),

\[
diag(Y) = [1 \ 1 \ldots 1]
\]  

(9)
Reformulate the problem to a convex formulation,
\[
\min_y (\delta + \beta \|Y\|^*)
\]
subject to
\[
Y \geq 0, \quad \text{diag}(Y) = [1 1 \ldots 1], \quad (10)
\]
where \( \beta \) is the Lagrange-multiplier.

The rank of \( A \) is the same as \( \Phi \) and \( S \). Here we proceed a factorization for matrix \( Y \) to obtain \( A \).
\[
Y = A^T A = USU^T, A = (Usqrt(S))^T \quad (11)
\]
Then we can get \( \Phi \) by calculating the pseudo inverse of \( \Psi \), which can be represented as:
\[
\Psi = USV^T, \quad \Psi^* = VS^{-1}U^T, \quad \Phi = A\Psi \quad (12)
\]

### 3.3 BINARY OPTIMAL PROJECTION

The derived sensing matrix is a float-point matrix, and in order to embed it into the sensor nodes to perform binary optimal projection (BOP), we employed a method to transform it to be binary \[14\]. The optimization problem can be formulated as
\[
\min_{Y,T} \|TY - \Phi\|_F^2
\]
subject to
\[
T^T T = I, \quad (13)
\]
\[
\Phi \in \{-1, 1\}^{m \times n}
\]
An heuristic solution can be obtained by considering each variable at one time with the other one assumed known. When \( \hat{\Phi} \) is known, optimal \( T \) can be found by
\[
T = UV^T. \quad (14)
\]
When \( T \) is known, the optimal \( \hat{\Phi} \) can be reached by simple quantization as
\[
\hat{\Psi}_{i,j} = \begin{cases} 1, & (T\Phi)_{i,j} \geq 0 \\ -1, & (T\Phi)_{i,j} < 0 \end{cases}
\]
The quantization of matrix \( \hat{\Psi} \) and the update of matrix \( T \) is processed iteratively until the error between the optimal and transformed binary sensing matrix is acceptable.

The feature of our sensing matrix design is that it is compatible with our trained dictionary. Dictionary learning ameliorates signal sparsity, thus the reconstruction quality. The update of sensing matrix also improves noise resistance capability, so under the same noise level, we can achieve the same recovery quality by less measurements compared to RS. The combination of both process enables the best recovery performance under noise with reduced measurements. Besides, the binary format of the sensing matrix is energy efficient, hence appropriate for hardware implementation.

### 3.4 DICTIONARY LEARNING

Overcomplete representations have been advocated because they have greater robustness in the presence of noise, can be more sparse, and can have greater flexibility in matching structure in the data. According to Candès, \( f \) is S-sparse under basis \( \Psi \), then if
\[
m \geq Cn\mu^2(\Phi, \Psi)S \log(n)
\]
for some positive constant \( C \), then we have an overwhelming probability to exactly reconstruct the original signal.

We seek the dictionary that leads to the best representation for every item in the dataset under sparsity constraints. The advantage of the trained dictionaries using machine learning techniques is they are fine tuned compared to the pre-determined ones, so their performance is significantly better. Here Online Dictionary Learning (ODL) is adopted. Compared to the methods mentioned above, ODL has higher training speed and requires less storage space because of its elimination of large matrix factorizations. With ODL, it is possible to add new features into the dictionary without stalling the reconstruction, which offers a mechanic of melioration when a distinctive input is received.

As each patient has its own biological characteristic, so for data acquired from different distributed sensors, the corresponding dictionary should be chosen for reconstruction. This can be done by sending the patient’s encoding at the very beginning before monitoring. This tells the server whom the data belongs to.

The dictionary learning part can be separated as sparse coding stage and dictionary update stage. The sparse coding problem is a \( l_1 \)-regularized least-squares problem defined as
\[
\alpha_t = \arg \min_{\alpha \in \mathbb{R}^n} \frac{1}{2} \|x_t - D_{t-1} \alpha\|_2^2 + \lambda \|\alpha\|_1. \quad (15)
\]
Due to the high correlations between columns of the dictionary, a Cholesky-based implementation of the LARS-Lasso algorithm, which provides the whole regularization path, is chosen here to solve the sparse coding problem \[8\].

At this stage, the objective is to find a dictionary \( D \) that satisfies:
\[
D_{t} = \arg \min_{D} \frac{1}{T} \sum_{i=1}^{T} \frac{1}{2} \|x_i - D \alpha_i\|_2^2 + \lambda \|\alpha_i\|_1. \quad (16)
\]
This problem can be solved by the block coordinate descent algorithm \[8\].

### 4. EXPERIMENTS

#### 4.1 Experiment setup

The compression ratio (CR) and reconstructed signal-noise-ratio (RSNR) are used as performance metrics.
\[
CR = n/m \quad (17)
\]
where \( n \) is the dimension of original signal \( x \), \( m \) is the
number of measurements.

\[
RSNR = \frac{\|x\|_2}{\|x - x'\|_2}
\]  

(18)

where \(x'\) is the reconstructed signal.

To evaluate our proposed binary optimal projection (BOP) and optimal projection (OP) framework, we compared their performances with the traditional random sensing (RS) framework with predetermined basis and trained basis.

We choose synthesized ECG dataset as our test bench [9]. The ECG sampling frequency is 128 Hz. The signal is segmented into 3000 samples for training and 1000 for testing, the dimension of each sample is 50 (n=50).

4.2 RSNR vs CR

We tested the reconstruction quality versus compression ratio (CR). Figure 2 illustrates the recovery quality (RSNR) versus CR under signal-noise-ratio (SNR) of 10dB and 50dB, respectively. It can be seen that our proposed CS framework, denoted by the yellow and green, outperforms random Bernoulli projections in terms of RSNR for all range of measurements \(m\) compared to traditional random Bernoulli sensing with DCT-DWT joint basis and trained dictionary. When SNR = 10dB, our proposed BOP and OP framework have an average 6dB and 9dB improvement in terms of RSNR compared to the RS framework with trained dictionary. When SNR = 50dB, the average improvement is 5dB and 8dB. Under the same noise level, OP and BOP has larger possibility to conform to RIP than RS, so the average recovery quality is superior to RS. The sensing matrix after binarization does not experience massive decrease compared to the optimal projection. As dictionary training improves the sparsity, and according to CS theory, we can use less measurements to achieve the same reconstruction quality. By fulfilling the requirement of RIP, the algorithm can tolerate more noise compared to the random projection. This means when noise is exposed, we can further reduce the data size required for wireless transmission.

4.3 RSNR vs SNR

We tested the reconstruction quality under different noise levels. For CS recovery under noise, we perform CS reconstruction via Basic Pursuit Denoising (BPDN) algorithm. Figure 3 plots the recovery quality versus SNR under CR of 5 and 9, respectively. It can be seen that the proposed framework outperforms random Bernoulli projections in terms of RSNR for all noise levels. When CR = 5, our proposed BOP and OP framework have an average 6dB and 9dB improvement in terms of RSNR compared to the RS framework with trained dictionary. When CR = 9, the average improvement is 2dB and 6dB. This solves the filtering problem faced by all traditional Nyquist sampling system. Because in practical applications, the random Gaussian
noise covers the frequency band of the signal in interest. So it is impossible to remove noise spanned in signal band. However in CS denoising theory, as the noise is not sparse under dictionary Ψ, so we can perfectly reconstruct the original signal as long as ΦΨ satisfies the RIP. We can see that the performance of our proposed system is stable for different SNR, this means it is not so sensitive to noise compared to RS.

4.4 Power estimation

Figure 4 demonstrates the measurements required for achieving a RSNR of 10dB under different SNR. We can see when SNR is relatively low (10dB to 20dB), the proposed BOP and OP system outperform RS significantly, utilizing half of the measurements to achieve the same RSNR, the data needed to be transmitted is also halved.

Toward practical monitoring application, we adopt an sampling system model with wireless transmitter to model the energy consumption for transmitting data of 1 hour. The ADC consumes 100 fJ/Sample, transmitter consumes 3 nJ/bit. Here we assume the sensed data is quantized to 8 bits, then the energy consumption for each framework is shown in Table I. The BOP and OP system consume 12 % and 11 % power of the traditional Nyquist sampling system, 60 % and 56 % power of the RS system.

Table 1: Power consumption comparison of several frameworks.

| Framework | Nyquist | RS | BOP | OP |
|-----------|---------|----|-----|----|
| ADC (mW)  | 0.046   | 0.046 | 0.046 | 0.046 |
| Tx (mW)   | 11.06   | 2.26 | 1.30 | 1.23 |
| Total P (mW) | 11.10 | 2.26 | 1.35 | 1.27 |
| Normalized P (%) | 100 | 20.33 | 12.13 | 11.48 |

5. CONCLUSION

In this paper, we introduced a data-driven CS framework for long-term health monitoring, where power efficiency is a major concern on sensor nodes. By leveraging the ECG signal features of individual patient, we proposed a general framework to train personalized sensing matrix and dictionary. Compared with generic sensing matrix and sparse bases that work for all patients, the trained personalized sensing matrix and the dictionary show great resilience to noise and irrelevant signals, which significantly improves the signal quality. The optimized quality ensures far less number of measurements and wireless transmissions on the sensor nodes, which leads to significantly enhanced power efficiency and battery life. In addition, our sensing matrix can be transformed into binary sensing matrix, which has simpler logic circuit implementation and also requires much less memory to store. This further improves the power efficiency on sensor nodes.

6. REFERENCES

[1] ABO-ZAHHAD, M., H. A., AND MOHAMED, A. Compression of ecg signal based on compressive sensing and the extraction of significant features. International Journal of Communications, Network and System Sciences 8 (2015), 97–117.
[2] Candès, E. J. Compressive sampling. Proceedings of the International Congress of Mathematicians. (2006).
[3] CHEN, F., CHANDRASKASAN, A., AND STOJANOVIC, V. Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors. IEEE Journal of Solid-State Circuits 47, 3 (March 2012), 744–756.
[4] DUARTE-CARVAJALINO, J., AND SAPIRO, G. Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization. Image Processing, IEEE Transactions on 18, 7 (July 2009), 1395–1408.
[5] Hegde, C., et al. Numax: A convex approach for learning near-isometric linear embeddings. Signal Processing, IEEE Transactions on 63, 22 (Nov 2015), 6109–6121.
[6] Lee, S., Luan, J., AND Chou, P. A new approach to compressing ecg signals with trained overcomplete dictionary. In EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth) (Nov 2014), pp. 83–86.
[7] LIN, Z., LU, C., AND LI, H. Optimized Projections for Compressed Sensing via Direct Mutual Coherence Minimization. ArXiv e-prints (Aug. 2015).
[8] MAIRAL, J., ET AL. Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res. 11 (Mar. 2010), 19–60.
[9] McSHARRY, P., ET AL. A dynamical model for generating synthetic electrocardiogram signals. Biomedical Engineering, IEEE Transactions on 50, 3 (March 2003), 289–294.
[10] OMITTED FOR BLIND REVIEW.
[11] Pare, G., JAANA, M., AND Sicotte, C. Systematic review of home telemonitoring for chronic diseases: The evidence base. Journal of the American Medical Informatics Association (2007).
[12] Polania, L., ET AL. Compressed sensing based method for ecg compression. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (May 2011), pp. 761–764.
[13] Ren, F., AND MARKOVIC, D. 18.5 a configurable 12-to-237ks/s 12.8mw sparse-approximation engine for mobile exg data aggregation. In IEEE International Solid State Circuits Conference (ISSCC) (Feb 2015), pp. 1–3.
[14] Wang, Y., ET AL. Optimizing boolean embedding matrix for compressive sensing in ram crossbar. In Low Power Electronics and Design (ISLPED), 2015 IEEE/ACM International Symposium on (July 2015), pp. 13–18.