Low-Complexity Compressed Analysis in Eigenspace with Limited Labeled Data for Real-Time Electrocardiography Telemonitoring

Kai-Chieh Hsu†, Bo-Hong Cho†, Ching-Yao Chou∗ and An-Yeu (Andy) Wu, Fellow, IEEE
†Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan (R.O.C)
∗Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan (R.O.C)
{b03901026, b03901022}@ntu.edu.tw, endpj@access.ee.ntu.edu.tw, andyw@ntu.edu.tw

Abstract—To achieve real-time electrocardiography (ECG) telemonitoring, one of the major obstacles to overcome is the scarce bandwidth. Compressed sensing (CS) has emerged as a promising technique to greatly compress the ECG signal with little computation. Furthermore, with edge-classification, the data rate can be reduced by transmitting abnormal ECG signals only. However, there are three main limitations: limited amount of labeled ECG data, tight battery constraint of edge devices and low response time requirement. Task-driven dictionary learning (TDDL) appears as an appropriate classifier to render low complexity and high generalization. Combining CS with TDDL directly (CA-N) will degrade classification and require higher complexity model. In this paper, we propose an eigenspace-aided compressed analysis (CA-E) integrating principal component analysis (PCA), CS and TDDL, sustaining not only light complexity but high performance under exiguous labeled ECG dataset. Simulation results show that CA-E reduces about 67% parameters, 76% training time, 87% inference time and has a smaller accuracy variance to the CA-N counterpart.

Index Terms—Compressed Analysis, Task-Driven Dictionary Learning, Compressed sensing, Real-Time ECG Telemonitoring

I. INTRODUCTION

As the aging population and the number of chronic diseases increase, patient-centered telemonitoring has become a vital research field so far [1]. Electrocardiography (ECG) signal is a good indicator of many cardiac diseases, so real-time ECG monitoring has been surveyed actively [2], [3]. However, to monitor ECG in real-time, we need to overcome the exceeding traffic of data and the scarcity of bandwidth resource. On the one hand, we can reduce the dimension of the ECG signal. Since ECG signal is of high sample rate and requires a great amount of bandwidth, compressed techniques including digital wavelet transform (DWT) [4] and compressed sensing (CS) [5], [6] are proposed. With negligible compressed performance difference, DWT manages to have more light-weight sensors relative to the DWT-based counterpart, achieving a 37.1% extension in the node lifetime [2]. Therefore, CS is more aligned with real-time ECG monitoring.

On the other hand, we can transmit the abnormal ECG signal only, which requires edge-computing [7] for classification. Since abnormal ECG signals are far less than normal ECG signals, edge-classification is a reasonable and valuable choice to greatly reduce the amount of transmitted data. Nonetheless, there are some limitations of edge-classification in ECG signal. First, the labeled ECG dataset is far smaller than the image dataset because ECG signal labeling is time-consuming and requires the effort of physicians. In addition, for ECG signals, the abnormal ones are more scarce than normal ones. Therefore, to resolve limited labeled dataset challenge, the classifier should be of great generalization. Secondly, tight battery constraint of edge devices and low response time requirement of real-time ECG monitoring demand the classifier should be of low complexity. For these limitations, task-driven dictionary learning (TDDL) [8] can be introduced as an effective model. TDDL can survive in limited labeled dataset since sparse feature extraction layer suggests high generalization [9]. In addition, TDDL has low complexity and few parameters via co-optimizing two-layer structure of a sparse feature extraction and a simple classifier.

To take the advantages of both CS and TDDL, one traditional framework is based on the reconstructed analysis (RA). However, RA is outperformed by compressed analysis (CA) since the reconstruction algorithm in RA is of very high complexity [3]. In CA, a naive way is to extract sparse representation directly from compressed ECG signal. However, the CS guarantee for good reconstruction is not enough to promote the required sparsity for classification. The naive CA (CA-N) trades the computational complexity off against acceptable classification performance. In this paper, we proposed an eigenspace-aided CA framework with a very light-weight classifier for better classification accuracy, compared to the naive way even for a limited labeled ECG dataset. Our main contributions are summarized as follows:

1. Our proposed CA-E reduces about 67% parameters, 76% training time and 87% inference time and has a smaller performance variance to CA-N counterpart.
2. Our proposed CA-E outperforms the conventional support vector machine and dense neural network by about 10% margin when the number of data is halved.
3. We propose eigenspace-aided CA to integrate with principal component analysis (PCA), CS and TDDL, and provide the source code† in the hope to help peer researchers.

†The toolbox can be downloaded at https://github.com/kevin71104/ECG-Telemonitoring/tree/master/Eigenspace-aided_Compressed_Analysis
The rest of paper is organized as follows. Section II provides the background of this paper and reviews some related works. Section III proposes our compressed analysis framework. Simulation results are discussed in section IV and the conclusion is drawn in section V.

II. BACKGROUND

A. Dictionary Learning

Consider a dataset of $n$ samples $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{N \times n}$. Classical unsupervised dictionary learning techniques [10], [11] minimize the following empirical risk

$$g_u(D) \triangleq \frac{1}{n} \sum_{i=1}^{n} \ell_u(x_i, D)$$  (1)

over $D \in \mathcal{D} \triangleq \{D \in \mathbb{R}^{N \times d} \mid \|d_j\|_2 \leq 1, \forall j = 1, \ldots, d\}$, where $d_j$ is the $j$th column of $D$. $d$ is the number of atoms used in the dictionary and the unsupervised loss function $\ell_u$ is defined as

$$\ell_u(x, D) = \min_{\alpha \in \mathbb{R}^d} \frac{1}{2} \|x - D\alpha\|^2 + \lambda_1\|\alpha\|_1$$  (2)

which is known as LASSO or basis pursuit and can be solved using OMP [12] or FISTA [13].

In addition to reconstruction, dictionary learning can be viewed as unsupervised sparse representation learning layer, where the sparse representation can be used for training classifier. In other words, consider each signal $x$ has a label $y$ and the sparse representation vector $\alpha^o(x, D)$ is defined as

$$\alpha^o(x, D) = \arg\min_{\alpha \in \mathbb{R}^d} \frac{1}{2} \|x - D\alpha\|^2 + \lambda_1\|\alpha\|_1$$  (3)

We can formulate the statistical risk of classification as

$$\min_{W \in \mathcal{W}} E_{y,x} \left[ \ell_s(y, W, \alpha^o(x, D)) \right] + \frac{\nu}{2} \|W\|^2$$  (4)

where $\ell_s$ is any convex loss function to evaluate the classification accuracy, $W$ is the classifier, $\mathcal{W}$ is a convex set and $\nu$ is the regularization parameter. In [8], the authors proposed task-driven dictionary learning (TDDL) to co-optimize both dictionary $D$ and $W$ simultaneously, namely,

$$\min_{D \in \mathcal{D}, W \in \mathcal{W}} E_{y,x} \left[ \ell_s(y, W, \alpha^o(x, D)) \right] + \frac{\nu}{2} \|W\|^2$$  (5)

Therefore, $D$ is learned in supervised vision and has enhance the performance dramatically.

B. Compressed Sensing

In CS framework, the original signal is sensed through a compressed random projection as

$$\hat{x} = \Phi x$$  (6)

where $\hat{x} \in \mathbb{R}^M$, $\Phi \in \mathbb{R}^{M \times N}$, $M < N$ and $M$ is the compressed dimension. We defined the compression ratio as $C_r \triangleq M/N$. To accurately reconstruct $x$ from $\hat{x}$, the $\Phi$ should be as incoherent to the sparse basis of $x$ as possible [6]. In addition, we can guarantee the incoherence above mentioned if the entries of $\Phi$ are chosen from identically independent distribution (i.i.d.) of Bernoulli(0.5) or Gaussian(0, 1) [6].

III. PROPOSED COMPRESSED ANALYSIS SCHEME

Since the dictionary and classifier should be independent of sensing matrix, we cannot directly apply TDDL on compressed ECG signal. We propose two compressed analysis: one is based on that CS guarantee perfect reconstruction if the sensing matrix is incoherent to the sparse basis; the other is based on PCA to express compressed ECG signal in the same basis. Below presents the details of two proposed methods. We use $X, \hat{x}, D, W$ and $\Phi$, defined in section II, to denote training dataset, testing data, dictionary, classifier and sensing matrix, respectively.

A. Naive Compressed Analysis

The $D$ and $W$ in naive compressed analysis (CA-N) are obtained via optimization of (5) using TDDL on $X$, where $W \in \mathbb{R}^{d \times N_c}$ is chosen as a simple linear regression and $N_c$ is the amount of classes.

If we can ensure the incoherence between $D$ and $\Phi$, we can formulate a new dictionary for compressed ECG testing signal of each sensing matrix as

$$\Xi \triangleq \Phi \cdot D$$  (7)

and the sparse representation vector can be obtained by

$$\alpha^o(\hat{x}, \Xi) = \arg\min_{\alpha \in \mathbb{R}^d} \frac{1}{2} \|\hat{x} - \Xi\alpha\|^2 + \lambda_1\|\alpha\|_1$$  (8)

where $\alpha^o(\hat{x}, \Xi)$ is dependent on $\Phi$. Once we obtained $\alpha^o(\hat{x}, \Xi)$, by multiplying it with $W$ and choosing the index of the largest entry, we can get the predicted class.

B. Eigenspace-aided Compressed Analysis

However, since the the target of the dictionary changes from reconstruction to feature extraction, the CS guarantee for good reconstruction is not enough to promise the similarity of the sparse representation vector obtained from (3) and (8).

We propose an eigenspace-aided compressed analysis (CA-E), utilizing PCA-assisted pre-processing to transform $\hat{x}$ from different $\Phi$ into the same subspace and followed by TDDL to not only enhance the classification performance and decrease the size of $D$.

To use PCA to obtain the eigenspace, we first apply singular value decomposition (SVD) on $X - \mu$, namely,

$$X' \triangleq X - \mu = U \cdot \Sigma \cdot V^T$$  (9)

where $\mu$ is the mean vector from each row of $X$, $U$ and $V$ are the left and right singular matrix and $\Sigma$ is a diagonal matrix whose diagonal entries are singular values sorted from largest to smallest. Since small singular values only stands for negligible part of representation or noise, we can only use the first $r$ largest singular values and extract signal subspace from corresponding singular vectors as

$$X_r = \Psi \cdot \Sigma_r \cdot V^T \triangleq \Psi \cdot T$$  (10)

where $\Psi$ and $\Sigma_r$ are the submatrices containing first $r$ columns of $U$ and rows of $\Sigma$, $X_r$ is the PCA-reconstruct
signal matrix and $T \in \mathbb{R}^{r \times n}$, the representation on $\Psi$, can be expressed as

$$T = \Psi^T \cdot X = \Sigma_r \cdot V^T$$  \hspace{1cm} (11)

Then, we apply TDDL on $T$ to get the $D$ and $W$.

As for testing data $\hat{x}$, Fig. 1 shows the signal flow of inference in CA-E. We first assume $x$ has below representation $x = \Psi s$. Therefore,

$$\hat{x}' = \hat{x} - \Phi \mu = \Phi \cdot \psi s = \Theta s$$  \hspace{1cm} (12)

where $\Theta = \Phi \cdot \Psi$. The representation vector $s$ is obtained by

$$s = \Theta^+ \hat{x}'$$  \hspace{1cm} (13)

where $\{\}^+$ denotes the pseudo-inverse. We transform $\hat{x}$ into the same subspace independent of sensing matrices and $\alpha^r(s, D)$ can be obtained by (3). Once we obtained $\alpha^r(s, D)$, multiplying it by $W$ and finding the index of biggest entry, we get the predicted class.

IV. SIMULATION RESULTS

A. Experimental Settings

We use a case study of atrial fibrillation (AF) detection to validate the benefits of our proposed algorithm. ECG signals are recorded in the intensive care unit (ICU) of stroke in National Taiwan University Hospital (NTUH). There are in total 231 non-AF records and 58 AF records labeled by doctors where each record spans ten minutes at the sampling frequency of 512 Hz. We randomly sample 2250 seconds and divide these samples into 1250 for training and 1000 for testing from non-AF and AF records, respectively. We first project ECG signals by a random matrix $\Phi$ to produce the compressed ECG signals as in (6). The entries of $\Phi$ are randomly chosen from $i.i.d. Bernoulli(0.5)$ and the dimension of $\Phi$ depends on $C_r$.

We compare our proposed eigenspace-aided CA (CA-E) with naive CA (CA-N) and other conventional classifiers, such as support vector machine (SVM) and dense neural network (DNN). In the implementation of CA-N and CA-E, we use FISTA [13] to obtain the sparse representation vector in (3) and (8). The classifier is chosen as a simple linear regression with one-hot encoding label and the loss function is chosen as $\ell_2$-loss. In addition, we initialize the dictionary $D$ by ODL [11], [14] and classifier $W$ by linear regression on sparse representation from $D$. In the CA-E, the number of eigenvectors used is set as $r = 83$. We decide hyper-parameters of models by validation set and the hyper-parameters pool are summarized in Table. I. In all scenarios, the number ratio, $N_r$, is defined as the amount of used training data divided by 1250 and 100 Monte Carlo simulations are executed on Python3 for each point.

B. Experimental Results

First of all, our proposed CA-E outperforms CA-N in the computational complexity under the similar classification performance. Fig. 2. shows the classification performance under different numbers of atoms in the dictionary. It is observed conventional methods of classification, DNN and SVM, have accuracy about 80 $\sim$ 85 percent. To surpass the conventional methods, CA-E merely requires 30 atoms in the dictionary, while CA-N needs 60 atoms. We compare CA-E with $d = 50$ (CA-E-50) and CA-N with $d = 100$ (CA-N-100) to demonstrate the training and inference overhead under the similar classification performance and CA-N with $d = 50$ (CA-N-50) is also chosen to illustrate the performance under the same number of atoms. Table. II. summarizes the comparison results. Under similar classification accuracy, CA-E-50 reduces about 67$\%$ parameters, 76$\%$ training time, 87$\%$ inference time and has a smaller performance variance relative to CA-N-100.

To elaborate on the complexity of training and inference overhead, we analyze the TDDL algorithm and find the bottleneck lies in FISTA, which is used to solve (3) and (8). Table. III. shows that CA-E-50 has about one-sixth FISTA time of CA-N-100 in training and one-eighth in inference. Since FISTA is a gradient-based optimization, the execution time of FISTA is based on the number of iterations used and the complexity of each iteration. The complexity of each iteration

![Fig. 1: The on-line inference mode of our proposed eigenspace-aided CA.](image1)

![Fig. 2: Classification accuracy under different number of atoms (d) with N_r = 0.6 and C_r = 0.25.](image2)

---

**TABLE I: Parameters setting for learning models**

<table>
<thead>
<tr>
<th></th>
<th>CA-E and CA-N</th>
<th>SVM</th>
<th>DNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell_2$-Constraint ($\lambda$)</td>
<td>$[0.2, 0.5, 0.8]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regularization ($\nu$)</td>
<td>$[10^{-4}, 10^{-7}]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kernel</td>
<td>Radial Basis Function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma ($\gamma$)</td>
<td>$[0.08, 0.10, 0.12, 0.15, 0.2]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost ($C_r$)</td>
<td>$[500, 800, 1000]$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hidden Layer Dimension</td>
<td>$[1, 16, 32, (32, 64), (64, 128), (128, 256), (8, 16, 32), (16, 32, 64), (32, 64, 128)]$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE II: Comparison table of CA-N and proposed CA-E with \( N_c = 2, C_r = 0.25, M = 128, r = 83 \) and \( N_r = 0.6 \).

<table>
<thead>
<tr>
<th>Model</th>
<th>The number of parameters</th>
<th>Training Time</th>
<th>Inference Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M \times d \times d \times N_c ) (6.5k)</td>
<td>183.82</td>
<td>106.32</td>
</tr>
<tr>
<td>CA-N (d = 50)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CA-N (d = 100)</td>
<td></td>
<td>452.56</td>
<td>306.33</td>
</tr>
<tr>
<td>CA-E (d = 50)</td>
<td>( r \times d \times d \times N_c ) (4.25k)</td>
<td>107.15</td>
<td>61.39</td>
</tr>
</tbody>
</table>

Training and testing time are measured on Intel i5-4200M CPU @ 2.5 GHz.

Fig. 3: Optimization process of FISTA in inference with \( C_r = 0.25 \).

Fig. 4: Classification accuracy under different data ratio \( (N_r) \) with \( C_r = 0.25 \).

depends heavily on the dimension of the dictionary, which is about \( O(d^2) \). Table II also illustrates that the number of iterations necessary in CA-E-50 is about two-fifths of in CA-N-100, which can also be seen in Fig. 3. The termination criterion is set as when \( \| \alpha^{t+1} - \alpha^t \|_2 \leq 1 \times 10^{-8} \) and \( t \) is the iteration index. Therefore, our proposed CA-E not only reduces the number of iterations needed in FISTA but also the complexity of each iteration of FISTA.

Secondly, our proposed CA-E is more immune to limited data challenge. Fig. 4 shows that CA-E has the best performance under most of the data ratios. It is also observed that all methods have accuracy over 80 percent under sufficient data \( (N_r \geq 0.9) \). Nevertheless, when lower data ratio is provided \( (N_r \sim 0.5) \), the performance of SVM and DNN dramatically drops below 80 percent, while CA-based algorithms still render classification accuracy over 80 percent. Hence, we can indicate that CA-based models have the feature of "high generalization" to survive under limited data challenge. Having a closer look into CA-based models, CA-E-50 outperforms CA-N-50 under all varieties of data ratio and performs slightly better CA-N-100. That is to say, our proposed CA-E-50 has high generalization like CA-based method does, with the computational complexity only about 20% of the CA-N-100.

Last but not least, to acquire the benefits of compress analysis, we have to address the entailed problems of variation of compression ratio. Fig. 5 demonstrates that our proposed CA-E has accuracy about 90 percent under all kinds of compression ratio, while other conventional models such as SVM and DNN have only about 80 percent. We can also inspect that, to reach the same level of performance to CA-E, we shall use 100 atoms in CA-N which is of higher complexity. As a consequence, CA-E is the lightest model to keep better classification performance under different compression ratios.

V. CONCLUSION

For real-time ECG telemonitoring with low complexity and high generalization, we propose an eigenspace-aided compressed analysis (CA-E). Simulation results demonstrate the advantages of CA-E. The number of needed iterations of FISTA in CA-E decreases, and meanwhile the complexity of each iteration is alleviated. Therefore, CA-E reduces about 67% parameters, 76% training time, 87% inference time and has a smaller performance variance relative to CA-N.

VI. ACKNOWLEDGEMENT

This research was supported in part by the Ministry of Science and Technology of Taiwan (MOST 106-2221-E-002-205-MY3, MOST 106-2622-8-002-013-TA, and MOST 107-2622-8-002-009-TA) and National Taiwan University. The first two authors contributed equally.
REFERENCES


463