Domain Adaptation for Online ECG Monitoring

Diego Carrera*, Beatrice Rossi†, Pasqualina Fragneto† and Giacomo Boracchi*

*Politecnico di Milano, Milano (MI) 20133, Italy
{diego.carrera,giacomo.boracchi}@polimi.it

†STMicroelectronics, Agrate Brianza (MB) 20864, Italy
{beatrice.rossi,pasqualina.fragneto}@st.com

***Abstract*—Successful ECG monitoring algorithms often rely on learned models to describe the heartbeats morphology. Unfortunately, when the heart rate increases the heartbeats get transformed, and a model that can properly describe the heartbeats of a specific user in resting conditions might not be appropriate for monitoring the same user during everyday activities. We model heartbeats by dictionaries yielding sparse representations and propose a novel domain-adaptation solution which transforms user-specific dictionaries according to the heart rate. In particular, we learn suitable linear transformations from a large dataset containing ECG tracings, and we show that these transformations can successfully adapt dictionaries when the heart rate changes. Remarkably, the same transformations can be used for multiple users and different sensing apparatus. We investigate the implications of our findings in ECG monitoring by wearable devices, and present an efficient implementation of an anomaly-detection algorithm leveraging such transformations.**

I. Introduction

Representation-learning algorithms [1], [2] play a central role in health and ECG monitoring, where they are employed to learn user-specific models describing the heartbeat morphology: many successful algorithms [3], [4], [5] suggest that these are necessary to achieve state-of-the-art performance in the detection/classification of dangerous arrhythmias. The next frontier is to directly perform ECG monitoring on small, low-power and wearable devices, thus facilitating the transition from hospital to houses. Despite wearable devices can be nowadays equipped with a suitable sensor suite and the computing capabilities to analyze ECG tracings while being acquired [6], using learned models during long-term monitoring of everyday activities is far from being straightforward. In particular, learned models have to be adapted to track heart rate variations, which is the problem we address here.

As shown in Figure 1, heartbeats get transformed when the heart rate changes, and they become different from the training set used to learn the user-specific model. Then, learned models might not properly describe heartbeats acquired during long recordings, and the ECG monitoring performance might degrade. Surprisingly, this problem has so far received very little attention, and most of monitoring solutions leveraging learned models have been tested on benchmark datasets containing short ECG tracings [7]. Straightforward approaches like, for instance, learning a model from extended training set including multiple heart rates for each user are not feasible, since heartbeats' morphology changes with electrodes location [5] and learned models refer specific device positioning. Moreover, to keep the training procedure simple and without risks, training data can be only acquired in resting conditions, without substantial variation in the heart rates.

Here we propose an appealing and feasible solution, which consists in adapting any user-specific model to track heart rate variations. This approach calls for the *domain-adaptation* or *transductive transfer learning* scenario [10], where a model learned from a *source domain* (heartbeats of a specific user in resting conditions), is modified / transformed to operate in a *target domain*, where observations are different (heartbeats of the same user at different heart rates). Our settings imply that training data for each user are available only in the source domain.

We consider dictionaries yielding sparse representations [2], [11], which are very effective to detect anomalies in ECG tracings [6], [12]. Domain adaptation techniques specifically designed for dictionaries go under the name of *dictionary adaptation*, and have been mainly developed in image classification [13], [14], [15]. These cannot be straightforwardly adapted to ECG monitoring scenario. In particular, [13] constrains data in source and target domains to share the same sparse representations, which is not appropriate for the many heartbeats composing an ECG tracing. In [14] dictionary adaptation is performed by a sequence of intermediate-domain dictionaries, which are learned from training samples in the source and target domains for each user. Unfortunately, training heartbeats for a specific user can be only acquired in resting conditions (source domain). The dictionary-adaptation framework in [15] projects both source and target data in a common subspace, and can be customized for ECG monitoring.

Our domain-adaptation solution maps user-specific dictionaries through a set of linear and user-independent transformations. Despite ECG signals heavily depend on the user, the sensing device and the electrodes position, different human hearts feature similar electrical conduction systems. For this reason, we use a large dataset of ECG tracings from several users [8] to learn transformations that adapt dictionaries as a function of the heart rate. We successfully learn these transformations by minimizing a functional that maps low-dimensional subspaces from source to target domain, requiring that transformed (source-domain) dictionaries can provide sparse representations of heartbeats at the target heart rate.

We state the addressed problem in Section II and describe the proposed solution in Section III, while in Section IV we show how to use the learned transformations to perform

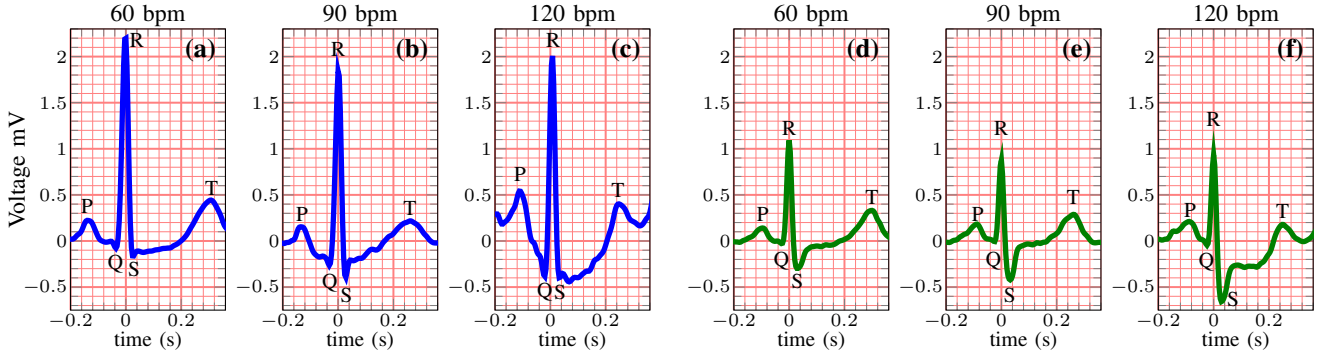

Figure 1. Examples of heartbeats recorded at different heart rates for user s20071 (**a**-**c**) and user s20431 (**d**-**f**) in [8]. Letters indicate the P-waves, the QRS-complexes and the T-waves [9], as they are referred to in the medical literature, while the heart rates are reported over each plot. Heartbeats in the **a** and **d** have been acquired at 60 bpm. When the heart rate increases (**b**, **c**, **e**, **f**), the supports of all heartbeats contract, the T-waves approach the QRS complexes and the QT intervals narrow down. Heartbeats of the two users exhibit a different morphology, which is also influenced by the electrodes location, However all the heartbeats undergo a similar transformation which does not seem to be a simple dilation / contraction, since peaks change their intensities and shapes.

online ECG monitoring. Considering the resource-constrained settings characterizing battery-powered wearable devices, we present in Section V an optimized implementation of the sparse-coding algorithm to substantially reduce the computational complexity of online ECG monitoring. Our experiments in Section VI demonstrate the effectiveness of the proposed solution on both heartbeat reconstruction and anomaly detection, and show that it can outperform state-of-the-art alternatives [15], as well as baseline solutions. Moreover, we show that transformations learned from a public dataset can successfully map dictionaries learned from a wearable device, equipped with a very different sensing apparatus. The conclusions along with future works are presented in Section VII.

II. Problem Formulation

Let us denote by $\mathbf{s}_{u,r} \in \mathbb{R}^{p(r)}$ an heartbeat acquired from the user u at the heart rate r, where the heartbeat support $p(r)$ depends on the heart rate. According to [6], we assume that heartbeats can be conveniently approximated by a *sparse representation* w.r.t. a dictionary $D_{u,r} \in \mathbb{R}^{p(r)\times n}$:

$$\mathbf{s}_{u,r} \approx D_{u,r}\mathbf{x}_{u,r}, \tag{1}$$

where the representation $\mathbf{x}_{u,r} \in \mathbb{R}^n$ has only few nonzero components, and n is the number of columns of $D_{u,r}$, which are referred to as the dictionary atoms. The user-specific dictionary $D_{u,r}$ is learned [11] from a training set $S_{u,r} \in \mathbb{R}^{p(r)\times m}$ containing m heartbeats of the user u at heart rate r. As in domain adaptation, training data are provided for each user u only in the source domain, namely at heart rate r_0.

The problem we address here is to analyze heartbeats of the same user acquired at different heart rates $r_1 \neq r_0$, i.e., in the target domain, where no training data are available to learn a dictionary D_{u,r_1}. To this purpose, we adapt the dictionary learned from the user u in the source domain, i.e. D_{u,r_0}, by means of a transformation $\mathcal{F}_{r_1,r_0}\colon \mathbb{R}^{p(r_0)\times n} \to \mathbb{R}^{p(r_1)\times n}$

$$\widehat{D}_{u,r_1} = \mathcal{F}_{r_1,r_0}(D_{u,r_0}) \quad \forall u, \tag{2}$$

and use the transformed dictionary $\widehat{D}_{u,r_1}$ to approximate the hearbeats of the user u in the target-domain. Our problem is formulated as learning a set of user-independent transformations $\{\mathcal{F}_{r_1,r_0}\}$, for multiple pairs (r_0, r_1), which transform the dictionary of any user at source heart rate r_0, to make it able to operate at target heart rate r_1.

III. Learning Transformations

Learning a dictionary for ECG monitoring corresponds to learning multiple, low-dimensional subspaces of $\mathbb{R}^{p(r)}$ where the heartbeats of each specific user live. Each of such subspaces is spanned by few atoms of the dictionary, and we assume in (1) that training heartbeats can be well approximated by a suitable union of low-dimensional subspaces. Sparse coding [16] corresponds to identifying the closest subspace, and then projecting $\mathbf{s}$ in there.

To preserve the structure of these low-dimensional subspaces, we assume transformations $\{\mathcal{F}_{r_1,r_0}\}$ to be linear. A general linear mapping from $\mathbb{R}^{p(r_0)\times n}$ to $\mathbb{R}^{p(r_1)\times n}$ has $n^2 p(r_0)p(r_1)$ degrees of freedom, which can be quite a huge number considering that heartbeats are typically composed of more than hundred of samples (in our datasets $p(r_0), p(r_1) \approx 150$). Therefore, we learn, for each pair (r_0, r_1), a linear transformation that maps each atom of D_{u,r_0} into an atom in the target domain, such that the union of low-dimensional subspaces generated by the transformed atoms is close to the heartbeats at heart rate r_1. This reduces the degrees of freedom of $\mathcal{F}_{r_1,r_0}$ to $p(r_0)p(r_1)$. The transformed dictionaries are obtained by multiplication by a matrix $F_{r_1,r_0} \in \mathbb{R}^{p(r_0)\times p(r_1)}$, as follows:

$$\widehat{D}_{u,r_1} = F_{r_1,r_0} D_{u,r_0} \qquad \forall u. \tag{3}$$

As discussed in Section I, transformations $\{\mathcal{F}_{r_1,r_0}\}$ should be user-independent, and we learn them from publicly available datasets [8] containing heartbeats of $L \gg 1$ users at different heart rates. In particular, for each pair (r_0, r_1), we learn the matrix F_{r_1,r_0} by solving the following optimization problem:

$$F_{r_1,r_0} = \underset{F,\{X_u\}_u}{\arg\min} \frac{1}{2}\sum_{u=1}^{L} \|S_{u,r_1} - F D_{u,r_0} X_u\|_2^2 + \mu \sum_{u=1}^{L} \|X_u\|_1 + \\ + \frac{\lambda}{2}\|W \odot F\|_2^2 + \xi \|W \odot F\|_1, \tag{4}$$

where S_{u,r_1} is a matrix stacking in its columns all the heartbeats from the user u at the heart rate r_1, and the columns of $X_u \in \mathbb{R}^{n\times m}$ contain the sparse representation of the corresponding heartbeats. Hadamard product is denoted by $\odot$.

The functional in (4) is composed as follows: the reconstruction error of the transformed dictionaries is the sum of $\|S_{u,r_1} - FD_{u,r_0}X_u\|_2^2$. Three regularization terms are controlled by parameters $\mu, \lambda, \xi \geq 0$ and a weight matrix $W \in \mathbb{R}^{p(r_1)\times p(r_0)}$. The first regularization term ensures that dictionaries FD_{u,r_0} provide good and sparse approximations of training heartbeats in $S_{u,r_1}, u \in \{1,\dots,L\}$. The sparsity is enforced by penalizing the ℓ^1 norm of the coefficient vectors. The other two terms represent a weighted elastic net penalization, which improves the overall stability of the learning process, and at the same time steers the estimated F_{r_1,r_0} towards desirable properties specified by W. In particular, we expect the transformation $\mathcal{F}_{r_1,r_0}$ to be local, namely that the output of each sample of transformed atoms in D_{u,r_1} is determined by few neighboring samples of the corresponding input atom. To this purpose, we set W to feature larger weights in positions far from the diagonal, as shown in Figure 2(a).

The problem (4) is not jointly convex in $\{X_u\}_u$ and F, but the functional to be minimized is convex with respect to each variable when the others are fixed. Therefore, we solve (4) by the Alternating Direction Method of Multipliers (ADMM), which enjoys good convergence properties in this case [17]. The rationale behind the ADMM is to split the optimization problem in many sub-problems, and alternate their optimization. To this purpose, we introduce the auxiliary variables $G \in \mathbb{R}^{p(r_1)\times p(r_0)}$ and $Y_u \in \mathbb{R}^{n\times m}$, $u \in \{1,\dots,L\}$, and (4) becomes equivalent to the following problem:

$$\begin{aligned}\underset{F,\{X_u\}_u,G,\{Y_u\}_u}{\arg\min}\ & \frac{1}{2}\sum_{u=1}^{L}\|S_{u,r_1} - FD_{u,r_0}X_u\|_2^2 + \mu\sum_{u=1}^{L}\|Y_u\|_1 + \\ & + \frac{\lambda}{2}\|W\odot G\|_2^2 + \xi\|W\odot G\|_1, \quad \text{s.t. } F = G,\ X_u = Y_u,\ \forall u.\end{aligned} \tag{5}$$

According to the ADMM framework, we define the Augmented Lagrangian of (5) and alternate the minimizations w.r.t. each variable while keeping the others fixed, and the update of Lagrange multipliers. The ADMM iterations are given by:

$$\begin{aligned}X_u^{(k+1)} &= \underset{X}{\arg\min}\ \frac{1}{2}\left\|S_{u,r_1} - F^{(k)}D_{u,r_0}X\right\|_2^2 + \\ &\quad + \frac{\rho}{2}\left\|X - Y_u^{(k)} + Z_u^{(k)}\right\|_2^2,\ \forall u \qquad (6)\\ Y_u^{(k+1)} &= \underset{Y}{\arg\min}\ \mu\|Y\|_1 + \frac{\rho}{2}\left\|X_u^{(k+1)} - Y + Z_u^{(k)}\right\|_2^2,\ \forall u \qquad (7)\\ F^{(k+1)} &= \underset{F}{\arg\min}\ \frac{1}{2}\sum_{u=1}^{L}\left\|S_{u,r_1} - FD_{u,r_0}X_u^{(k+1)}\right\|_2^2 + \\ &\quad \frac{\sigma}{2}\left\|F - G^{(k)} + H^{(k)}\right\|_2^2, \qquad (8)\\ G^{(k+1)} &= \underset{G}{\arg\min}\ \frac{\lambda}{2}\|W\odot G\|_2^2 + \xi\|W\odot G\|_1 + \\ &\quad \frac{\sigma}{2}\left\|F^{(k+1)} - G + H^{(k)}\right\|_2^2, \qquad (9)\\ Z_u^{(k+1)} &= Z_u^{(k)} + X_u^{(k+1)} - Y_u^{(k+1)},\ \forall u \qquad (10)\\ H^{(k+1)} &= H^{(k)} + F^{(k+1)} - G^{(k+1)}, \qquad (11)\end{aligned}$$

where $Z_u \in \mathbb{R}^{n\times m}$, $u \in \{1,\dots,L\}$, and $H \in \mathbb{R}^{p(r_1)\times p(r_0)}$ are the Lagrange multipliers of the constraints in (5).

Sub-problems (6) and (8) are standard quadratic expressions which can be efficiently solved by a linear system, using techniques such as Gaussian elimination. Problem (7) admits a closed-form solution as the proximal mapping [18] of $\mu/\rho||\cdot||_1$

$$[Y_u^{(k+1)}]_{ij} = \mathcal{S}_{\mu/\rho}\Big(\big[X_u^{(k+1)} + Z_u^{(k)}\big]_{ij}\Big),$$

where $[\cdot]_{ij}$ denotes the entry of the i-th row and the j-th column of a matrix, and $\mathcal{S}_\eta\colon \mathbb{R}\to\mathbb{R}$ is the soft-thresholding operator defined as $\mathcal{S}_\eta(x) = \text{sign}(x)\cdot\max\{0, x-\eta\}$. It can be shown by calculus and [18] that (9) can be also solved by the soft thresholding as:

$$[G^{(k+1)}]_{ij} = \frac{1}{1+\lambda[W]_{ij}^2/\sigma}\mathcal{S}_{\xi[W]_{ij}/\sigma}\left(\left[F^{(k+1)} + H^{(k)}\right]_{ij}\right).$$

We initialize the ADMM algorithm by setting to zero the values of all the variables but $F^{(0)}$, which is initialized to uniformly distributed random values to avoid trivial solutions. Then, we iteratively solve (6)-(11) until a maximum number of iterations is met.

An example of learned F_{r_1,r_0} is shown in Figure 2(b), where we see that the nonzero elements are concentrated around the diagonal, thanks to W in the elastic net regularization. The central rows have smaller support than the first and last rows, indicating that the QRS complexes are less influenced by the transformation than the P and T waves, as shown in Figure 1.

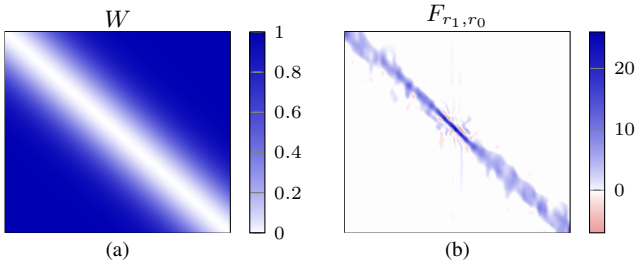

Figure 2. (a) Weight matrix W used to learn transformations. The element (i,j) of W is defined as $[W]_{ij} = 1 - c\cdot\exp(-(j-i)^2/\omega)$, where ω determines the width of the Gaussian and the constant c is set to ensure that the value of weights is between 0 and 1. Setting larger weights far from the diagonal is a way to promote locality of the learned transformation. (b) Example of learned $\widehat{F}_{r_1,r_0}$ for $r_1 = 90$, $r_0 = 70$ using the weight matrix W shown in (a). The learned transformation is local, since nonzero elements of $\widehat{F}_{r_1,r_0}$ are concentrated around the diagonal.

IV. Online ECG Monitoring

Here we describe how the proposed domain-adaptation solution can be used to perform online monitoring of ECG tracings directly on a wearable device. In particular, we focus on the detection of anomalous heartbeats, such as arrhythmias. For an effective processing during long-term monitoring, the transformations $\{\mathcal{F}_{r_1,r_0}\}$, estimated as described in Section III, are stored on the wearable device and the heart rates (r_0, r_1) are quantized in a range having resolution of 5 beats per minute (bpm).

ECG Preprocessing. We remove the baseline wander by standard preprocessing techniques [19]. Then, we extract heartbeats by cropping a window of $p(r)$ samples around each R-peak, which are detected by the Pan-Tompkins algorithm [10]. R-peaks are used also to estimate the reference heart rate r of each heartbeat, by averaging the reciprocals of the R-R intervals over a window of 5 seconds.

User-configuration. During the device configuration on each user u, we record 10 minutes of ECG tracing in resting conditions and collect the heartbeats in the training set S_{u,r_0}. This is used to learn the user-specific dictionary D_{u,r_0} by KSVD algorithm [2], where the source heart rate r_0 is the most frequent heart rate in the collected tracing. Then, we compute $\widehat{D}_{u,r_1} = F_{r_1,r_0} D_{u,r_0}$ for each target heart rate r_1, and store in memory all dictionaries $\{\widehat{D}_{u,r_1}\}$.

Domain Adaptation. During online monitoring, we estimate the current heart rate r_1, and select $\widehat{D}_{u,r_1}$ among the stored dictionaries.

Anomaly Detection. Anomalous heartbeats are detected as in [6], identifying as anomalous any heartbeat $\mathbf{s}_{u,r_1}$ that is not sufficiently close to the union of low-dimensional subspaces generated by $\widehat{D}_{u,r_1}$. To this purpose, the first step consists in computing the sparse representation $\widehat{\mathbf{x}}_{u,r_1}$ of $\mathbf{s}_{u,r_1}$ w.r.t. to $\widehat{D}_{u,r_1}$, by solving the *sparse coding* problem

$$\widehat{\mathbf{x}}_{u,r_1} = \arg\min_{\mathbf{x}} \left\| \mathbf{s}_{u,r_1} - \widehat{D}_{u,r_1}\mathbf{x} \right\|_2^2, \text{ s.t. } \|\mathbf{x}\|_0 \leq \kappa, \quad (12)$$

where the functional to be minimized is the ℓ^2 norm of the residual w.r.t. to the dictionary, and $\|\mathbf{x}\|_0$ is the number of nonzero coefficients in $\mathbf{x}$. Sparse coding (12) is an NP-Hard problem, and we compute a suboptimal solution through a greedy algorithm, the Orthogonal Matching Pursuit (OMP) [20]. The reconstruction error $e(\mathbf{s}_{u,r_1}) = \|\mathbf{s}_{u,r_1} - \widehat{D}_{u,r_1}\widehat{\mathbf{x}}_{u,r_1}\|_2$ can be used as an indicator to discriminate between normal heartbeats (that conform with $\widehat{D}_{u,r_1}$) and anomalous heartbeats (that do not conform with $\widehat{D}_{u,r_1}$). In particular, we adopt the rule:

$$\mathbf{s}_{u,r_1} \text{ is anomalous} \quad \Leftrightarrow \quad e(\mathbf{s}_{u,r_1}) > \gamma, \quad (13)$$

which recognizes as anomalous any heartbeat that cannot be properly reconstructed by the transformed dictionary $\widehat{D}_{u,r_1}$. The threshold $\gamma > 0$ determines the promptness of the detector to identify anomalous heartbeats.

V. Optimized Processing for Wearable Devices

The study in [6] shows that dictionaries $\widehat{D} \in \mathbb{R}^{p\times n}$ having much less atoms than the heartbeat dimension, i.e. $n \ll p$, are more effective in detecting ECG anomalies. This implies that the union of low dimensional subspaces where normal heartbeats live can be enclosed in a n-dimensional subspace of $\mathbb{R}^p$. Since $\mathbf{s}$ has to be projected into such n-dimensional subspace, it is possible to reduce the number of operations performed in the OMP by changing the basis of the subspace generated by the columns of $\widehat{D}$. This coordinate change lead to replacing (12) by the following sparse-coding problem, where we have omitted the subscript (u, r_1), for the notation sake:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\| Q^T\mathbf{s} - R\mathbf{x} \right\|_2^2, \text{ s.t. } \|\mathbf{x}\|_0 \leq \kappa. \quad (14)$$

In (14), $\widehat{D} = QR$ is the QR decomposition of $\widehat{D}$, $R \in \mathbb{R}^{n\times n}$ is a square and upper-triangular matrix, and $Q \in \mathbb{R}^{p\times n}$ is a rectangular matrix whose columns are orthonormal and span the subspace generated by the atoms of D, such that $Q^TQ = I_n$. The following proposition shows that problem (14) and (12) has the same solution.

Proposition 1. *For every* $\mathbf{s} \in \mathbb{R}^p$ *and* $\mathbf{x} \in \mathbb{R}^n$ *it holds:*

$$\left\| \mathbf{s} - \widehat{D}\mathbf{x} \right\|_2^2 = \left\| Q^T\mathbf{s} - R\mathbf{x} \right\|_2^2 + \|\mathbf{s}\|_2^2 - \left\| Q^T\mathbf{s} \right\|_2^2. \quad (15)$$

Proof follows from $Q^TQ = I_n$ and basic linear algebra. Solving (12) is much less expensive than (14), since the computational complexity of OMP is determined by the target sparsity κ and the dictionary size[1] [20]. When solving (12), the dictionary $\widehat{D}$ has size $p \times n$ and the complexity is $O(\kappa pn)$, since in our case $\kappa < n \ll p$. In contrast, since $R \in \mathbb{R}^{n\times n}$, solving (14) requires only a complexity of $O(\kappa n^2)$ on top of the cost of computing of $Q^T\mathbf{s}$, which is $O(pn)$ and becomes dominant. In practice, we have reduced the overall cost of OMP from $O(\kappa pn)$ to $O(pn)$ by first transforming each heartbeat $\mathbf{s}$ as $Q^T\mathbf{s}$, and then computing the OMP w.r.t. R.

It is convenient to modify the user-configuration phase by computing the QR decomposition of each transformed dictionary $F_{r_1,r_0} D_{u,r_0} = \widehat{D}_{u,r_1} = Q_{u,r_1} R_{u,r_1}$, and storing the matrices Q_{u,r_1} and R_{u,r_1} in place of $\widehat{D}_{u,r_1}$ to perform online monitoring. Sparse coding has to be solved for each acquired heartbeat, thus the proposed solution, which is more efficient than traditional one, can meaningfully increase the battery life.

VI. Experiments

Our experiments are meant to demonstrate that the learned transformations can be successfully used to perform domain adaptation *i*) on multiple users, and *ii*) on users monitored though different devices. In particular, we assess adaptation performance by signal reconstruction and anomaly-detection experiments, and we implement a leave-one-out procedure to prove that transformations learned from a large public dataset [8], can effectively adapt dictionaries of multiple users. We also show that the same transformations can be successfully used on the *Bio2Bit Move dataset*, which was acquired by a completely different sensing device.

A. Datasets Description

The *Long-Term ST dataset* (LT-ST) [8] contains long term ECG recordings from 80 users, lasting from 21 to 24 hours and is manually annotated by experts. We consider only two types of heartbeats: normal, which are used for learning the transformations, and anomalous, which include all the annotated arrhythmias and that are used to test anomaly-detection performance. We consider heart rates $r \in \{70, 75, \ldots, 120\}$.

The *Bio2Bit Move Dataset* (B2B) contains 7 ECG tracings recorded from healthy users acquired by the Bio2Bit Move device [21], a prototypal wearable device developed by STMicroelectronics. Each ECG tracing lasts more than 1 hour and is acquired during normal-life activities (e.g. resting, lying down, walking, resting / walking after an effort): most of the heart rates refer to $r_1 \in \{80, 90, 100\}$. Due to the small distance between electrodes, heartbeats in the B2B dataset are very different from those in the LT-ST dataset.

[1]The exact number of floating point operations required by the OMP algorithm to solve (12) is: $2\kappa pn + 2pn + 2\kappa^2 p + 3\kappa p + 2p + 2\kappa n + \kappa^3 + \frac{3}{2}\kappa^2 + \frac{\kappa}{2}$.

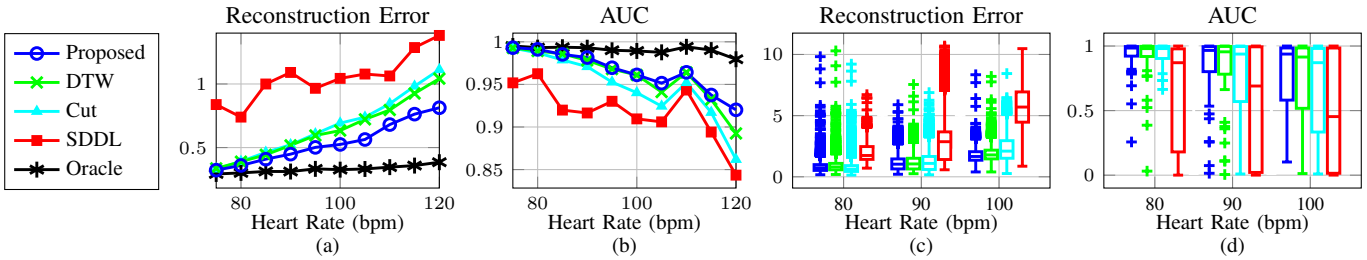

Figure 3. Results on LT-ST dataset (a-b) and B2B dataset (c-d). (a) The median reconstruction error over all users. As expected, in case of Oracle solution the reconstruction error is nearly constant w.r.t. to the heart rate, while it increases more significantly with r_1 for all the other domain adaptation solutions, confirming that the change in the heartbeat morphology becomes more evident for large heart rates. The proposed solution achieve the lowest reconstruction error. (b) The median AUC for the considered anomaly detection problem computed over all the users. The proposed solution leads to the best domain adaptation, although DTW achieves similar performance. (c) The boxplots of the reconstruction error computed over all the user in the datasets for each domain adaptation solutions. (d) The boxplots of the AUC achieved by all solutions to detect *inter-user* anomalous heartbeats. The proposed solution achieves the best performance according to both figures of merit, especially in case of $r_1 = 100$.

B. Figures of Merit

We consider two tasks: signal reconstruction and anomaly detection. Reconstruction error $e(\mathbf{s}_{u,r_1})$ indicates how good the transformed dictionaries $\widehat{D}_{u,r_1}$ are at modeling normal heartbeats in the target domain. The anomaly-detection performance indicates the ability to distinguish heartbeats featuring a morphology that is different from normal ones. As figures of merit we consider the True Positive Rate (TPR) and the False Positive Rate (FPR). Since both FPR and TPR depend on the threshold γ in (13), we compute the Receiving Operating Characteristic (ROC) curve by considering FPR and TPR for several values of γ, and use the Area Under the ROC Curve (AUC) as a global indicator.

C. Considered Solutions

We compare the proposed domain-adaptation solution with the following ones.

Cut: Figure 1 suggests that the support of the heartbeats contracts as the heart rate increase: thus the simplest form of dictionary adaptation consists in transformations $\mathcal{F}_{r_1,r_0}$ that remove the first and the last samples in each column of D_{u,r_0}.

DTW: we use dynamic time-warping [22], a classic algorithm to align vectors and measure their similarity, to compute the transformation matrix F_{r_1,r_0}. For each pair of heartbeats $\mathbf{s}_{r_0} \in \mathbb{R}^{p(r_0)}$ and $\mathbf{s}_{r_1} \in \mathbb{R}^{p(r_1)}$, dynamic time-warping computes a non uniform resampling of $\mathbf{s}_{r_0}$ and $\mathbf{s}_{r_1}$ to obtain two aligned vectors $\widetilde{\mathbf{s}}_{r_0}, \widetilde{\mathbf{s}}_{r_1} \in \mathbb{R}^{\widetilde{p}}$ having a common (and larger) support $\widetilde{p} \geq p(r_0), p(r_1)$. These resamplings can be expressed as $\widetilde{\mathbf{s}}_{r_i} = A_i \mathbf{s}_{r_i}$, $i = 0, 1$, and are computed by minimizing the euclidean distance between the aligned vectors. We estimate the matrices $A_i \in \mathbb{R}^{\widetilde{p} \times p(r_i)}$ by aligning the first principal components of S_{u,r_0} and S_{u,r_1}. To obtain user-independent transformations, we compute matrices A_i by minimizing the sum of Euclidean distances between the aligned principal components over all the users. Transformations $\{\mathcal{F}_{r_1,r_0}\}$ are defined by setting $F_{r_1,r_0} = A_{r_1}^{+} A_{r_0}$ in (3), where $A_{r_1}^{+}$ denotes the pseudo-inverse of A_{r_1}.

SDDL: this solution is obtained by adapting Shared Domain-adaptive Dictionary Learning solution presented [15] to our settings. The solution in [15] jointly learns two projections from the source and target domains onto a common subspace together with a shared dictionary that provides sparse representations of all the projected data. In SDDL we learn the projections for each pair (r_0, r_1) from LT-ST dataset. Then, we perform user-configuration by *i*) projecting the training source heartbeats at heart rate r_0 onto the low-dimensional subspace, and *ii*) learning a user-specific dictionary by means of the KSVD [2]. During online monitoring, we project each heartbeat at heart rate r_1 onto the low-dimensional subspaces using the learned projection, compute the sparse representation w.r.t. to the user-specific dictionary, and back-project the obtained reconstruction to the target domain. The reconstruction error is then the ℓ^2 norm of the difference between the original heartbeat and the back-projected one.

Oracle: this is an ideal solution, which cannot be pursued in practice as it uses a training set in each target domain. Oracle represents an upper bound on the performance of the considered domain-adaptation solutions. In particular, we do not estimate any transformation $\mathcal{F}_{r_1,r_0}$, but directly a dictionary D_{u,r_1} from the training data S_{u,r_1} using KSVD.

D. Experiments on LT-ST Dataset

We learn from the LT-ST dataset the transformations $\{\mathcal{F}_{r_1,r_0}\}$ for the proposed and DTW solutions, as well as the projection matrices of SDDL. Our goal is to assess whether the learned transformations / projections can be used to adapt dictionaries learned from users that were not in the training set. To this purpose, we implement a leave-one-out procedure where each test set contains heartbeats of a single user. As in nested cross-validation, the training users are analyzed in a 5-fold cross-validation scheme to learn the transformations / projections and tune hyper-parameters. In particular, we use random search [23] to select hyper-parameters that minimize the average reconstruction error of the heartbeats of the validation users in the target domain. Heartbeats of the test user are solely used for performance assessment.

Figure 3(a) shows that the median of the average reconstruction errors over the users (the lower the better) steadily increases with the heart rates and that the proposed solution outperforms all the others, in particular of large r_1. A Mann-Whitney U test confirms that when $r_1 \geq 80$ the average reconstruction error achieved by the proposed solution is lower

than that of the DTW (p-value < 0.001), which is the second-best performing solution. When $r_1 = 75$ there is no clear statistical evidence (p-value $= 0.02$), since the morphologies of source and target heartbeats are similar and the Cut solution is also quite successful. In contrast, as r_1 increases, more flexible transformations are needed to perform domain-adaptation. Figure 3(b) reports the median AUC (the higher the better) at different heart rates, and that the best domain-adaptation solution is the proposed one. DTW achieves similar performance, and the Mann-Whitney U test confirms that there is no significant differences between the two solutions. Thus, anomalous heartbeats are better perceivable at low heart rates, and dictionaries adapted using the baseline solution perform comparably to ours even though they achieve lower reconstruction performance. When the heart rate increases, it is important to properly adapt the dictionaries to discriminate normal and anomalous heartbeats. Not surprisingly, the Oracle solution outperforms all the others and maintains a nearly constant reconstruction error.

E. Experiments on B2B Dataset

We test whether transformations $\{\mathcal{F}_{r_1,r_0}\}$ learned on the LT-ST dataset can successfully adapt dictionaries learned from the B2B dataset. We learn a dictionary D_{u,r_0} for each of the 7 users and transform them to operate at $r_1 \in \{80, 90, 100\}$. Transformed dictionaries are used as in previous heartbeats-reconstruction and anomaly-detection experiments. The Oracle solution has not been considered due to the lack of heartbeats to properly learn the dictionary at some heart rates r_1.

The box-plots of the reconstruction error in Figure 3(c) are consistent with the previous results, indicating that the proposed solution outperforms the others. To assess the anomaly-detection performance, we proceed as in [6]: since these tracings do not contain arrhythmias, we consider as anomalies all the heartbeats coming from users that were not in the training set. These artificially introduced *inter-user* anomalies are good examples of heartbeats featuring a different morphology. Boxplots in Figure 3(d) indicate that the proposed solution is the one yielding the best AUC, and that learned transformations can be successfully adapt dictionaries on different devices.

VII. Conclusions

We address the domain-adaptation problem for dictionaries yielding sparse representations of ECG tracings. We propose a novel solution which learns user-independent transformations from publicly available datasets, and then map user-specific dictionaries to model heartbeats acquired at target heart rates. Our experiments show that learned transformations can be successfully applied on different users, confirming the intuition that heartbeat variations due to heart rate changes are similar for different users, even when their heartbeats feature different morphologies. Most importantly, we show that the learned transformations can be used to successfully adapt dictionaries learned on different sensing devices. This has relevant implications in wearable device scenarios, for which we present an optimized implementation of OMP. Ongoing work goes in the direction of developing further solutions for online ECG monitoring on wearable devices, including techniques to automatically discard heartbeats corrupted by motion artifacts. Moreover, we are investigating the use of our solution in other scenarios where no training data in the target domain are available, such as anomaly detection in images.

References

[1] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 8, pp. 1798–1828, 2013.

[2] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans Signal Process*, vol. 54, no. 11, pp. 4311–4322, 2006.

[3] J. Wiens and J. V. Guttag, "Active learning applied to patient-adaptive heartbeat classification," in *NIPS*, pp. 2442–2450, 2010.

[4] P. de Chazal and R. B. Reilly, "A patient-adapting heartbeat classifier using ecg morphology and heartbeat interval features," *IEEE Trans Biomed Eng*, vol. 53, no. 12, pp. 2535–2543, 2006.

[5] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ecg beat classifier using a mixture of experts approach," *IEEE Trans Biomed Eng*, vol. 44, no. 9, pp. 891–900, 1997.

[6] D. Carrera, B. Rossi, D. Zambon, P. Fragneto, and G. Boracchi, "Ecg monitoring in wearable devices by sparse models," in *ECML-PKDD*, pp. 145–160, 2016.

[7] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Eng Med Biol Mag*, vol. 20, no. 3, pp. 45–50, 2001.

[8] F. Jager, A. Taddei, G. B. Moody, M. Emdin, G. Antolič, R. Dorn, A. Smrdel, C. Marchesi, and R. G. Mark, "Long-term st database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia," *Med Biol Eng Comput*, vol. 41, no. 2, pp. 172–182, 2003.

[9] D. L. Mann, D. P. Zipes, P. Libby, and R. O. Bonow, *Braunwald's heart disease: a textbook of cardiovascular medicine*. Elsevier Health Sciences, 2014.

[10] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans Knowl Data Eng*, vol. 22, no. 10, pp. 1345–1359, 2010.

[11] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *ICML*, pp. 689–696, 2009.

[12] A. Adler, M. Elad, Y. Hel-Or, and E. Rivlin, "Sparse coding with anomaly detection," in *MLSP*, pp. 1–6, 2013.

[13] Q. Qiu, V. M. Patel, P. Turaga, and R. Chellappa, "Domain adaptive dictionary learning," in *ECCV*, pp. 631–645, 2012.

[14] J. Ni, Q. Qiu, and R. Chellappa, "Subspace interpolation via dictionary learning for unsupervised domain adaptation," in *CVPR*, pp. 692–699, 2013.

[15] S. Shekhar, V. M. Patel, H. V. Nguyen, and R. Chellappa, "Generalized domain-adaptive dictionaries," in *CVPR*, pp. 361–368, 2013.

[16] A. Coates and A. Y. Ng, "The importance of encoding versus training with sparse coding and vector quantization," in *ICML*, pp. 921–928, 2011.

[17] Q. Liu, S. Wang, J. Luo, Y. Zhu, and M. Ye, "An augmented lagrangian approach to general dictionary learning for image denoising," *J Vis Commun Image Represent*, vol. 23, no. 5, pp. 753–766, 2012.

[18] Y.-L. Yu, "On decomposing the proximal map," in *NIPS*, pp. 91–99, 2013.

[19] P. de Chazal, M. O. Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *IEEE Trans Biomed Eng*, vol. 51, no. 7, pp. 1196–1206, 2004.

[20] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," 2008.

[21] P. Gentile, M. Pessione, S. Antonio, Z. Alessandro, and I. Fernanda, "Embedded wearable integrating real-time processing of electromyography signals," in *Eurosensors*, 2017. Accepted.

[22] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans Acoust*, vol. 26, no. 1, pp. 43–49, 1978.

[23] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *NIPS*, pp. 2546–2554, 2011.