Patient-Friendly Detection of Early Peripheral Arterial Diseases (PAD) by Budgeted Sensor Selection

Qiaojun Wang
Department of Electrical and Computer Engineering
Rutgers University
Piscataway, NJ
Email: qjwang@eden.rutgers.edu

Kai Zhang
Siemens Corporation
Corporate Research and Technology
755 College Road East
Princeton, NJ, 08540
Email: kai-zhang@siemens.com

Ivan Marsic
Department of Electrical and Computer Engineering
Rutgers University
Piscataway, NJ
Email: marsic@ece.rutgers.edu

John K.J. Li
Department of Biomedical Engineering
Rutgers University
Piscataway, NJ
Email: johnkjli@rci.rutgers.edu

Fabian Moerchen
Siemens Corporation
Corporate Research and Technology
755 College Road East
Princeton, NJ, 08540
Email: fabian.moerchen@siemens.com

***Abstract*—Sensor networks provide a concise picture of complex systems and have been widely applied in health care domain. One typical scenario is to deploy sensors at different locations of human body and analyze the sensor measurements collectively to perform diagnosis of diseases. In this work, we are interested in differentiating peripheral arterial disease (PAD) patients from healthy people by monitoring peripheral blood pressure waveforms using electric sensors. PAD is an important cause of heart disease, which causes no significant symptoms until in a late stage. Therefore its early detection is of significant clinical values. Currently, PAD diagnosis either require large equipment or complicated, invasive sensor deployment, which is highly undesired in terms of medical expenses and safety considerations.**

To solve this problem, we present a novel approach to address the issue of high deployment cost in PAD detection via sensor networks. Assuming we are given many possibilities for sensor placement, each with different deployment cost, our goal is to select a small number of sensors with minimal costs while delivering accurate diagnosis. We solve this problem by treating each sensor as a feature, and designing a budget-constrained feature selection scheme to choose a compact, optimal subset of sensors, inducing very low deployment cost in terms of invasive treatment, while giving competitive classification accuracy compared with state-of-the-art feature selection method.

I. Introduction

Sensor networks have experienced significant developments in modern scientific and engineering domains in recent years. It can provide a concise picture of large, complex systems via aggregating a collection of nodes/sensors that capture parameters of interest, which renders great convenience and flexibility in a variety of applications involving system state classification. We are interested in applications of sensor networks approach in the health care domain. In health care, an intelligent sensor network can be used to automatically evaluate the health status of the subject and to react when some certain state of the system has been identified (such as danger, emergency health condition, or certain disease is diagnosed) [1], [2], [3] [4] [5]. In these applications, typically, a set of sensors are deployed on human body, and the sensor measurements are analyzed collectively through machine learning algorithms to perform classification. Advances of sensor technology and various data analysis and information processing algorithms make this a highly promising area. However, an important, practical concern in this scenario is that sensor deployment might involve highly invasive treatment on the human body, which is undesirable in terms of both medical expenses and safety considerations.

Consider the example of peripheral arterial disease (PAD) classification. The PAD affects millions of people in the United Stated, in which plaque builds up in the arteries and limits the flow of oxygen-rich blood to organs, see Figure 1. Usually patients have no obvious symptoms until in a late stage, therefore early diagnosis of PAD is very important in controlling the disease and reduce its damage. Currently, there are several approaches for PAD diagnosis but they have certain limitations. For example, Doppler Ultrasound or Magnetic Resonance Angiogram require expensive medical devices which might not be widely available in small clinics; Blood test requires taking the blood sample and performing chemical analysis; the Ankle-Brachial Index method needs to test the speed of blood flows which is usually complex and quite time consuming.

On the other hand, peripheral pressure waveforms have been recognized as an important factor in the evaluation of human arterial system integrity [6]. Therefore, we propose to deploy sensors on the human body to monitor the peripheral

PervasiveHealth 2012, May 21-24, San Diego, United States

DOI 10.4108/icst.pervasivehealth.2012.249068

pressure waveforms, which can then be analyzed through machine learning algorithms to identify the state of the arterial system. The sensor networks have much lower cost than large medical equipments and are more flexible to deploy, however a practical concern is that only less than 20% of the arteries are non-invasively measurable from body surface. The remaining arteries require different levels of invasiveness for conducting measurements. For example, if an artery is close to heart, the invasiveness would be very high at this artery, in terms of both medical costs and safety considerations. Therefore, how to faithfully detect the PAD by measuring as few arteries as possible, and as non-invasively as possible, is an important research topic of significant clinical values.

In this paper, we will focus on addressing the issue of high sensor deployment cost in PAD detection via sensor networks. The problem statement is formalized as follows. Assume we are given many possibilities for sensor placement or installation, each with different conditions (different locations of human body) and hence induces different cost in their deployment, and our goal is to select a small set of sensor positions which will induce low deployment cost while at the same time faithfully reflect the system condition and allow highly accurate classification based on the sensor readouts.

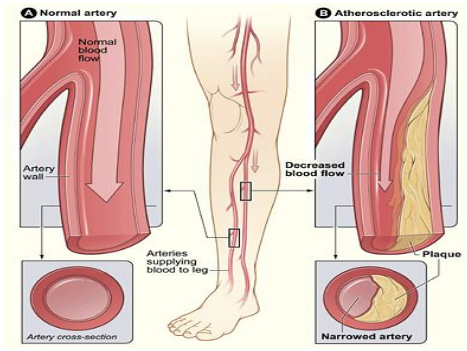

Fig. 1. Normal Artery versus narrowed artery in PAD. Picture by virtue of National Hear Lung and Blood Institute (http://www.nhlbi.nih.gov/health/health-topics/topics/pad/).

Our problem falls into the general problem of *sensor selection*, aimed at selecting an optimal set of sensors from a sensor network to optimize performance. Selection criteria varies a lot depending on domains. For example, in environmental applications *s*patial sampling design is used to make measurements of dynamic spatial processes [7][8]; some works minimize the error in estimating the parameters via convex programming [9], information theoretic framework [10], or geometric approaches [11]. These work do not consider factors related to the deployment cost of the sensors. Another class of methods considers reducing energy consumption of network by routing or topology control [12] or by utility-based sensor selection [13]. Again, their problem setting is quite different from ours. In [14], the authors considered the problem of sensor selection when redundancy relationships between sensors can be formulated through information network modeling. However, the resultant integer programming is very expensive and a greedy approximation algorithm is designed in practice.

Considering either the computational inefficiency and the distinct problem setting of existing methods, in this paper we propose a new approach to solve our problem. To achieve this, we introduce the concept of (supervised) feature selection in the domain of networks: the measurements collected by each node in the sensor network is deemed as a feature, and our objective is to select a small subset of informative features/nodes on which to build an accurate classifier. Feature selection is an effective tool to pick highly compact and informative subset of features for accurate classification. More importantly, a novel contribution of our work is to extend feature selection to a budget orientated framework: the features are selected not only to boost the classification performance, but also to enforce an effective control on the expected cost in deploying the related sensors. Our formulation builds upon and generalizes the maximum relevance minimum redundancy criterion (MRMR) [15]. Besides, we employ a soft, probabilistic decision scheme which allows the problem to be formulated neatly as a quadratic programming (QP) with a unique, globally optimal solution.

To test the efficiency of our proposed method, we applied it in differentiating PAD patients from healthy people by monitoring peripheral blood pressure waveforms using both invasive and non-invasive sensors at multiple locations. Compared with existing feature selection approaches, our method can select a small number of locations for sensor placement which induces very low deployment cost in terms of invasive treatment, while at the same time giving rise to more accurate classification results. Our approach provides a valuable guidance on early diagnosis of PAD.

The rest of this paper is organized as follows. In Section 2 we provide a global picture of our approach for PAD detection, including Data acquisition (Section 2.1), SVM classification (Section 2.2), and Feature Selection (Section 2.3). In Section 3, we review existing feature selection methods. In Section 4, we propose our soft, budgeted feature selection method based on the MRMR criteria. In Section 5 we compare our approach with state-of-the-art feature selection algorithm in the task of PAD detection. The last section concludes the paper.

II. PAD Detection Using Sensor Networks

In this section, we provide a global picture of our PAD detection scheme, including data acquisition (Section2.1), SVM classification (Section 2.2), and feature selection (Section 2.3). The basic idea is to deploy a sensor network at a number of arteries in human body, each keeping record of the peripheral pressure waveforms for that artery. Our objective is then to detect whether the subject suffers from PAD by analyzing the measurements from the sensor network. We deem this as a classification problem, and solve it by support vector machine (SVM). In order to improve the classification performance, reduce the number of sensors needed and the associated cost of measurement invasiveness, we will apply a budgeted feature selection algorithm, which will be the main technical contribution of the paper and whose discussion will be deferred in

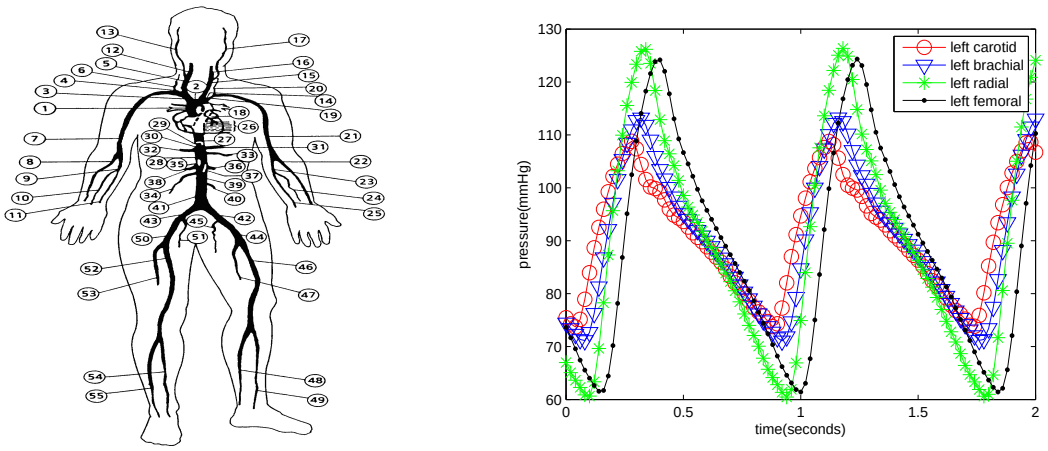

Fig. 2. The human arterial system illustration (left) and the predicted pressure waveforms using the model in [31] (right).

Section 4.

A. Data Acquisition

Acquisition of peripheral pressure waveforms is difficult, because the price of both invasive sensors and the tonometer-based noninvasive monitors is high. There is almost no publicly available data of peripheral pressure waveforms at all arteries in human body. On the other hand, at current technical levels, the accuracy of such sensors still remains unstable. Currently, we are working closely with electric sensor engineers and clinical doctors to improve the sensor measurement accuracy. As a first step in conducting our research, we therefore resort to model based simulation. Recent advances in vascular biology and vascular engineering have led to the understanding and integration of the two fields [16]. In this paper, we use a distributed multi-branching arterial tree for modeling the phenomenon of blood pressure waveform propagation and reflection developed in [17].

The construction of the arterial tree model consists of three steps: single segment modeling, interaction between segments, and multi-segment network connections. The basic computational entity is a segment of artery which is a thin-walled cylindrical tube having internal viscous, elastic and inertial properties with external coupling to the surrounding tissue which produces a longitudinal constraint. Let p_1 and p_2 be pressures of the source and sink points of the blood flow for a single segment. Then the following relations hold:

$$p_1 = A_1 e^{jwt}, \;\; p_2 = p_1 e^{-\tau l} = A_1 e^{jwt} e^{-\tau l},$$

where w is the frequency, l is the length of the segment, and τ is the propagation constant. The model also simulates the wave reflections which occur from any discontinuity along the arterial tree (e.g. branching points, areas of alteration in arterial distensibility, and high resistance terminal beds), and the amplitude and phase changes at reflection sites are modeled as $p_r = p_f \Gamma$, where Γ is reflecting coefficient. Finally, the whole arterial system of a human body is translated into a 55-segment arterial model shown schematically in Fig.1(a).

The predicted pressure waveform using the model [17] contains important information about the properties of arterial wall that is embedded in the propagation and reflection characteristics. The model was evaluated in terms of branch reflection coefficient, terminal vascular bed behavior, and wall viscoelasticity [17]. Figure 2(right) provides example waveforms simulated by this model at certain locations in the human body. Researchers have found that the model-predicted pressure waveforms compared favorably with real blood pressure waveforms, therefore, it serves as a faithful source of data generation and as we shall see, our analysis based on these data will provide important guidance on human arterial behavior. In Section 5, we will discuss in more detail how waveforms of healthy and PAD patients are generated by tuning a set of physiological parameters in the model.

B. Classification Using SVM

With data obtained from our model, we perform PAD detection as a classification problem. In the training phase, we collect measurements from a number of subjects, labeled either as positive (PAD) or negative (healthy), and train a classifier. In the testing phase, given a new subject we then apply the trained classifier to predict his/her status.

We used support vector machine in our experiment. The SVM [18] is state-of-the-art classification algorithm, and has been applied in many domains such as bioinformatics [19] and sensor networks [20]. Given a training samples $\mathbf{x}_i$'s and their labels y_i's (either 1 or -1), SVM computes the following decision function,

$$f(\mathbf{x}) = \sum_i \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b,$$

by maximizing the margin between the positive and negative classes, where α_i's are the coefficients returned by the optimization process. Here $K(\cdot, \cdot)$ is the kernel function that maps the input data $\mathbf{x}$ to a reproducing kernel Hilbert space, which allows us to solve linearly non-separable problems conveniently. Given any new testing sample $\mathbf{x}$, if $f(\mathbf{x}) \geq 0$, then $\mathbf{x}$ will be categorized as positive class (PAD patient);

otherwise $\mathbf{x}$ will be categorized as negative class (healthy people).

C. Feature Selection

The electric sensors used to measure the peripheral pulse pressures have to be placed on important positions of human body. In clinical settings, more than 80% of these measurements have to be conducted invasively, which is very undesirable considering both medical expenses and safety issues. Therefore, an important objective of our work is to faithfully perform PAD detection , *using as few sensors as possible, and with minimal invasiveness*. To achieve this goal, we will use the technique of supervised *feature selection*. More specifically, each node in the sensor network is considered as a feature, and each subject (or sample) is represented by multiple features/sensors. Given a set of training samples and their labels (either a healthy person or a PAD patient) the goal of feature selection is to select the most representative set of features based on which a classifier can be built and new testing samples can be classified accurately.

The importance of feature selection is three-fold. First, it can improve the prediction performance by removing noisy or irrelevant features, as verified in bioinformatics applications where the number of features typically exceeds thousands. In our problem, it is quite likely that not all the waveforms are useful for diagnosis and some sensors could record data that are either redundant or less informative which should be removed for classification purpose. Second, feature selection leads to faster testing, by reducing data dimensionality. In our problem this means less sensors are needed when measuring a new subject, which is desirable from practical views. Third, feature selection can sometimes give a better understanding of the underlying complex system process that generated the data [21].

Traditional feature selection focuses on the classification performance but does not take into account the cost of acquiring the features. In our application of PAD detection, however, acquisition of different features (via sensors at different locations of human body) will involve different cost depending on the level of invasiveness. Obviously, using invasive sensors for measuring arterial waveforms is highly undesirable. Therefore, standard feature selection methods are not directly suited in this problem. Our objective therefore is to select a subset of sensors (features) that are non-invasive or minimally invasive, while being able to deliver required classification accuracy. To achieve this goal, we extend current feature selection method to a budget oriented version, which while performing supervised feature selection simultaneously controls the cost associated with obtaining the selected features. We will discuss it in Section 4.

The usefulness of feature selection will be most pronounced during the testing phase: when a new subject arrives, we can choose to measure the waveforms using only a few, highly non-invasive sensors based on our budgeted feature selection results. In the training phase, however, note that we still need most (if not all) features to be able to perform feature selection. Once the training stage is done, the knowledge and model obtained can be used for highly user-friendly testing/diagnosis.

III. Feature Selection Review

In this section we briefly review existing feature selection methods. The *filter methods* for feature selection use simple statistics of individual features as a ranking score, such as correlation coefficient [22], and fisher score [23]. This is efficient but ignores correlation between features. In contrast, *wrapper methods* use the output of a classifier to assess the relative usefulness of subsets of features [24], such as the SVM method on recursive feature elimination [19]. The wrapper method repeats the training/testing many times, which can be computationally very expensive.

MRMR Criterion: Recently, the maximum relevance and minimum dependency (MRMR) feature selection method was proposed [15], which presents state-of-the-art result in gene expression data analysis. The idea is to choose a subset of features that: (i) have the minimum redundancy among themselves; and (ii) have maximum relevance with the target variables (class labels). The dependencies between features and class labels are measured by mutual information $I(\cdot,\cdot)$. Given a set of features F_i's for $i = 1, 2, ..., d$, the label y, and let S be the subset of features to be selected, the objective is

$$\max \frac{1}{|S|} \sum_{F_i \in S} I(F_i, y) - \frac{1}{|S|^2} \sum_{F_i, F_j \in S} I(F_i, F_j).$$

The first term is the relevance between selected features and the label/target (maximized); the second term is the total similarity among selected features (minimized). Though the MRMR criterion has demonstrated a lot of success, it has some limitations. First, it does not consider the cost of obtaining the features, which is an important concern in PAD detection. Second, it is solved by a greedy scheme which leads to sub-optimal solution that is dependent on initialization. At last, computing the mutual information for continuous random variables (which is needed in our problem) is still open problem, which requires either thresholding that leads to loss of information, or probability function estimation which itself is a challenging task. In the next section, we will propose a soft, budgeted feature selection method based on the MRMR criterion to address these concerns.

IV. Budgeted Node/Feature Selection in Sensor Networks

In this section, we propose a soft, budgeted feature selection method using a global optimization framework based on the MRMR criterion and budget constraint. There are two main contributions. First, our method directly incorporates the cost of feature acquisitions and enforces effective cost control, which is very suited for our application. Second, instead of using a greedy procedure for optimization as in [15], we employ a soft selection scheme, i.e., optimized for our problem is the probability that each feature should be selected. The relaxation to a soft probabilistic decision scheme allows us to formulate our problem as a quadratic programming (QP) for which globally optimal solution can be obtained.

A. Formulations

Suppose we have input data $\mathbf{x}_i \in \mathbb{R}^{d\times 1}$, for $i = 1, 2, ..., n$, where d is the number of features, and $\mathbf{y} \in \{\pm 1\}^{n\times 1}$ is the label. Let $F_j \in \mathbf{R}^{n\times 1}$ denote the jth feature for all training samples. Define $\mathbf{Q} \in \mathbb{R}^{d\times d}$ where Q_{ij} is the (non-negative) similarity between the ith and the jth feature, i.e., $Q_{ij} = sim(F_i, F_j)$. Let the cost associated with each feature be $c_i \geq 0$, $i = 1, 2, ..., d$, and let b be the overall budget for obtaining the features. Let $r_i \geq 0$ be the relevance between the ith feature and the target $\mathbf{y}$. We use p_i to represent the probability to select the ith feature, $i = 1, 2, ..., d$, and compute p_i's using the following optimization

$$\begin{aligned} \min_{\mathbf{p}\in\mathbf{R}^{d\times 1}} \quad & \mathbf{p}'\mathbf{Q}\mathbf{p} - \lambda \mathbf{r}'\mathbf{p} \\ \text{s.t.} \quad & p_i \geq 0 \\ & \textstyle\sum_{i=1}^{d} p_i = 1. \\ & \textstyle\sum_{i=1}^{d} c_i p_i \leq b \end{aligned} \tag{1}$$

The constraint $p_i \geq 0$ and $\sum_{i=1}^{d} p_i = 1$ are used to guarantee that $\{p_i, i = 1, 2, ..., d\}$ is a valid probability distribution. With this constraint, the first term $\mathbf{p}'\mathbf{Q}\mathbf{p} = \sum_{i=1}^{d}\sum_{j=1}^{d} p_i p_j Q_{ij}$ can then be deemed as the expectation of the similarity between selected features. This term is minimized to reduce the redundancy between selected features. The second term $\mathbf{r}'\mathbf{p} = \sum_i r_i p_i$ is the expectation of the relevance between selected features and the label. This term should be maximized to guarantee that selected features are useful for classification. Here λ is a positive regularization parameter to balance the strength of the two competing terms.

The constraint $\sum_{i=1}^{d} c_i p_i \leq b$ is used to control the cost of obtaining the features. Note that $\sum_{i=1}^{d} c_i p_i$ is the expected value of the cost of obtaining the features, therefore we require that it is below a pre-determined threshold budget b. Here, the budget b (as well as the cost c_i's) is expected to be domain specific and needs to be determined by the domain experts.

B. Feature Similarities

In this section we discuss more details on how to compute the similarity between features/targets. We have used the squared correlation coefficients

$$\mathbf{Q}_{ij} = (F_i^\top F_j)^2. \tag{2}$$

One can also normalize the vectors F_i's to have zero mean and unit variance, i.e.,

$$F_i^{\mathcal{N}} = \frac{F_i - \bar{F}_i}{\sqrt{var(F_i)}},$$

where $\bar{F}_i$ and $\sqrt{var(F_i)}$ is the mean and standard deviation of the entries in the vector F_i. One can also compute the relevance between a feature F_i and the target $\mathbf{y}$ as

$$r_i = (F_i^\top \mathbf{y})^2,$$

or use the normalized version.

The similarity measure (2) is exact and easy to compute compared with mutual information. On the other hand, the resultant Hessian matrix $\mathbf{Q}$ in (1) will be positive semi-definite (PSD). To see this, note that Hessian matrix there can be deemed as computing a polynomial kernel matrix using F_j's as "points" with degree 2. It is well known that any polynomial kernel is positive semi-definite [25]. Note that any Quadratic Programming (QP) problem with PSD Hessian matrix is guaranteed to have a globally optimal solution [26]. Therefore our problem (1) will have a global solution.

C. Postprocessing

Once the probabilities p_i's are computed, the corresponding features can then be selected based on the magnitudes of the probabilities. One sort p_i's and sequentially select features with large p_i's until a pre-specified *real budget* has been met. Here the real budget is different from the expected budget (8). The expected budget is the weighted sum of the costs of all the features, and it is bounded by the minimum and maximum cost of individual features. In comparison, the *real budget* is the actual sum of the costs of selected features. In practice, the actual budget can be chosen based on domain knowledge.

D. Quadratic Programming Feature Selection

Our probabilistic way of feature selection is similar to the quadratic programming feature selection framework proposed in [27]. However, a significant difference is that we have a budge constraint on the cost of selected features. In particular, considering that $p_i \geq 0$, the budget constraint $\sum c_i p_i \leq= \sum c_i |p_i| \leq b$ is equivalent to an adaptive L_1-norm regularization that enforces cost-dependent sparsity on the feature selection result.

In case there is a large number of features, the QP problem becomes prohibitively expensive, and [27] used Nyström method to approximate the Hessian matrix. Instead of performing random sampling as in the standard Nyström method, one can use some more advanced sampling scheme, such as the K-means based sampling proposed in [28] [29], which usually produces more accurate approximation given the same sampling rate.

V. Experiments

A. Experimental Setting

In this section, we report empirical results on early PAD detection, using the simulated waveform data generated by the arterial system model (section 2). Note that there are altogether 55 segments in the model, each involving 7 physiological parameters (Table 1), and hence there are altogether 55× 7 = 385 parameters needed to generate the data. The control values of the parameters (corresponding to a healthy person) can be found in ([6]) and in Table 1 we list as an example the parameter values for "radial artery" (segment-8). In Table 2, we list the 55 segments which are categorized into 3 groups with different costs, corresponding to non-invasive, low invasive, and highly invasive. The order of the segments is the same as marked in Figure 2(left).

We simulate a population of healthy people as follows. Using the control values of physiological parameters, we

add a random noise with magnitude around $\pm 10\%$ of the control value to simulate variations among individuals. These parameters are then fed to the simulator to generate the blood-pressure (BP) waveform for healthy people. For PAD patient, we consider altogether 55 sub-groups corresponding to the occlusion of each of the 55 segments. Each PAD segment occlusion is simulated by decreasing the radius from the control value. We used the following occlusion degrees: $\approx 10\%, \approx 20\%$ and $\approx 30\%$, which are considered low levels of PAD. In the simulation, this variability corresponds to a random noise subtracted from the control value of the artery radius. Again, the parameter can be used to simulate the BP waveform of a PAD patient with problems on any of the 55 segments. Note that based on the time series data generated by the simulator, we can compute the pulse pressure (the difference between the maximum and minimum pressures) at each of the 55 segments in the arterial tree (Figure 2), which will be used as our input features. Considering the periodic nature of the waveform, one can also extract more sophisticate features regarding the dynamics of the time series [30]. The task of feature selection is then to select the optimal subset of features (i.e. sensors) with minimum cost for PAD detection.

We have simulated the waveform data of 2000 people in total. Among them 50% of the samples are healthy people; the other 50% consist of a uniformly random selection of 55 types of PAD (corresponding to the 55 arterial segments). We randomly chose 60% of the data as the training data and the other 40% as testing data. We first apply the feature selection method to select a subset of k sensors based on the training data and the training labels; then we train a SVM classifier based on the training data using only the selected sensors; lastly, we perform prediction on the testing data, again only using the selected k sensors.

In our experiments we have used the RBF kernel in the SVM, which is well known to be able to map the data into an infinite-dimensional Hilbert space and allows us to learn a highly nonlinear decision function to model complex practical problems. We use libsvm[1] which is a state-of-the-art nonlinear SVM solver that is very efficient on medium-sized data sets. The parameters of the SVM are chosen as follows. The regularization parameter C is chosen as $\{0.1, 1, 10, 100, 1000, 10000\}$; the kernel width in the kernel function $k(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2\gamma\right)$ is chosen from $\gamma_0 \times \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 1, 2, 4, 8, 16, 32\}$, where γ_0 is chosen as the reciprocal of the averaged squared pairwise distances between the training instances, i.e., $\gamma_0^{-1} = \frac{1}{n}\sum_{i,j=1}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|^2$. Then we use 5-fold cross validation to choose the best candidate parameter in SVM, as well as λ in (5) which is the most widely used way of parameter selection in practical applications.

B. Results and Discussion

We compare the performance of the following methods: (1) Our method: the probabilistic budgeted feature selection method proposed in this paper; (2) MRMR: maximum relevance and minimum redundancy feature selection method proposed in [15]; (3) Random: feature selection of a random subset of features.

To perform a comprehensive evaluation of the feature selection capacity of different methods, we gradually increase the number of sensors they select in the range $[1, 20]$. We do not consider sensors for all 55 segments, because that is not a realistic scenario. When a subset of sensors is selected, we then use it to perform training and testing, and plot the accuracy-versus-cost curves for comparison. For our method, the computed probabilities p_i's provide a natural ranking of the importance of different sensors, therefore gradually increasing the number of sensors one by one using this order. For MRMR and random method, we can directly specify the number of sensors as an input to the method. We used the MRMR implementation from the authors' website[2]. Note that the MRMR uses the feature maximally aligned to the target as the initial start; therefore, the reported results are deterministic. In case a random initial feature is selected, the MRMR feature selection results would have more variations. For the random feature selection method, given a desired number of sensors to be selected, we repeat the experiments 100 times and compute the mean cost and mean accuracy, together with the standard deviation of the accuracies.

We also separate the simulated data into three categories according to the severity of occlusion used to create them, which includes three levels, $\approx 10\%, \approx 20\%$, and $\approx 30\%$ corresponding to slight, moderate, and medium PAD disease, and examine how different feature selection methods perform on them. We plot the results in Figure 3. As seen, as the number of selected sensors increases, the performance of all the three methods tend to improve. For our method and the MRMR method, the improvement of performance is significant when the number of sensors is very low, and the improvement slows down when the number of sensors or the total cost reach a certain threshold, indicating that both methods will select most informative features first, and additional features are more and more redundant. In comparison, for the random method, the performance improves linearly. In addition, it can be observed that patients with higher severity are generally easier to detect due to the more pronounced patterns in the input waveforms than those with lower severities.

The MRMR method demonstrates a significant improvement compared with the random feature selection, which validates the usefulness of the MRMR criterion as well as the fact that selecting informative sensors are quite beneficial to good classification performance. However, we note that MRMR does not take into account the cost of the sensors. Therefore although the sensors selected by MRMR can lead to a good accuracy, the cost of the selected sensors can also be quite large. In comparison, our method directly controls the cost of the selected sensors. As seen, with the same cost of measurement acquisition, our method almost always

[1] http://www.csie.ntu.edu.tw/∼cjlin/libsvm

[2] http://penglab.janelia.org/proj/mRMR/

leads to a better classification performance. In addition, to achieve a similar classification accuracy, our method will need lower cost. For both methods, the performance is significantly better than for random feature selection. Another interesting observation is that for patients with lower PAD severity (and hence more difficult detection problem), the classification improvement of our method compared with either MRMR or random method is more obvious. This indicates that our feature selection method is more suited for early PAD diagnosis.

In Figure 4, we illustrate the probabilities p_i's computed by our method by plotting them versus sensor deployment costs c_i's. As can be seen, Low-cost nodes typically have higher probabilities to be selected than high cost nodes[3], which clearly validates the budget control effect of our feature selection scheme. In particular, the many p_i's for the high cost sensors (half of all features) are exactly zeros, showing the sparsity of our feature selection results. This is consistent with the discussion in Section 4.2 on the connection of our approach with adaptive sparse modeling.

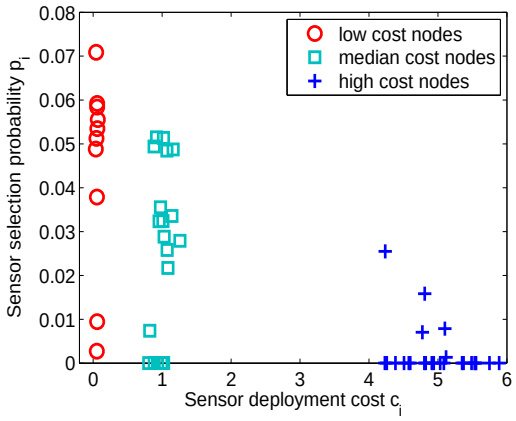

Fig. 4. The cost of each sensor (x-axis) and the probability of selecting the sensor (y-axis) computed by our method. Observations: (1) low-cost nodes are more likely to have a higher probability to be selected than high-cost nodes, reflecting the validity of budget constraint; (2) many probabilities (about 50%) are exactly zero, indicating a sparse feature selection result.

VI. Conclusions and Future Work

In this paper, we propose a budgeted sensor selection method to address the issue of high sensor deployment cost in system state classification via sensor networks, and apply it successfully in early PAD detection. Our method explicitly controls the cost of sensor deployment while preserving high accuracy of diagnosis, which has significant practical value in health care applications; it gives satisfactory accuracy in classifying healthy people and PAD patients, while at the same time requiring much lower cost of invasiveness compared with state-of-the-art feature selection approach.

[3]It is worthwhile to note that this does not necessarily mean lower cost nodes are intrinsically more useful, since our probabilities are computed as a combined effect of maximizing feature relevance and minimizing feature costs. It may be properly read as "given equally relevant features, our approach would favor the cheaper ones, thereby assigning higher probabilities to them".

Our study provides a promising way for early PAD diagnosis. The selected sensor positions can be used as an important guidance for physicians to select easy and patient-friendly sensor measurements to facilitate their decision. Now we are working closely with biomedical engineers on designing robust and reliable sensors and testing our approach on real data. In the future we will also be tackling the more challenging problem of PAD localization, i.e. identifying the location of PAD arteries, which is a more challenging classification problem with structured output. We will also apply our approach to other kind of networks with larger number of nodes [27].

References

[1] G. Krassnig, D. Tantinger, C. Hofmann, T. Wittenberg, and M. Struck, "User-friendly system for recognition of activities with an accelerometer," in *In PervasiveHealth*, 03 2010.

[2] B. Longstaff, S. Reddy, and D. Estrin, "Improving activity classification for health applications on mobile devices using active and semi-supervised learning," in *In PervasiveHealth*, 03 2010.

[3] O. Chipara, C. Lu, T. Bailey, and G.-C. Roman, "Reliable clinical monitoring using wireless sensor networks: Experience in a step-down hospital unit," in *ACM Conference on Embedded Networked Sensor Systems (SenSys)*, 2010.

[4] D. Jung, T. Teixeira, and A. Savvides, "Towards cooperative localization of wearable sensors using accelerometers and cameras," in *IEEE Infocom*, 2010.

[5] Z. Zeng, S. Yu, W. Shin, and J. C. Hou, "Pas: A wirelessenabled, cell-phone-incorporated personal assistant system for independent and assisted living," in *International Conference on Distributed Computing Systems*, 2008.

[6] H. Zhang, "A novel frequency domain arterial tree model: Distributed wave reflections and potential clinical applications," Ph.D. dissertation, Rutgers, the State University of New Jersey, 2006.

[7] Z. Zhu and M. L. Stein, "Spatial sampling design for prediction with estimated parameters," *Journal of Agricultural Biological and Environmental Statistics*, vol. 11, pp. 24–44, 2006.

[8] R. Olea, "sampling design optimization for spatial functions," *Mathematical Geology*, 1984.

[9] S. Joshi and S. Boyd, "Sensor selection via convex programming," *IEEE Transactions on Signal Processing*, pp. 451 – 462, Feb 2009.

[10] M. Chu, H. Haussecker, F. Zhao, M. Chu, H. Haussecker, and F. Zhao, "Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks," *International Journal of High Performance Computing Applications*, vol. 16, 2002.

[11] C. Giraud and B. Jouvencel, "Sensor selection: a geometric approach," in *IEEE/RSJ International Conference on Human Robot Interaction and Cooperative Robots*, 1995.

[12] Q. Dong, "maximizing system lifetime in wireless sensor networks," in *Fourth International Symposium on Information Processing in Sensor Networks.*, 2005.

[13] F. Bian, D. Kempe, and R. Govindan, "utility-based sensor selection," in *The Fifth International Conference on Information Processing in Sensor Networks.*, 2006.

[14] C. Aggarwal, A. Bar-Noy, and S. Shamoun, "On sensor selection in linked information networks," in *International Conference on Distributed Computing in Sensor Systems*, 06 2011.

[15] C. Huang, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, August 2005.

[16] J. K.-J. Li, *Dynamics of the Vascular System*. Singapore: World Scientific, 2004.

[17] H. Zhang and J.-J. Li, "A novel wave reflection model of the human arterial system," *Cardiovascular Engineering*, vol. 9, no. 2, pp. 39–48, 2009.

[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[19] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–442, 2002.

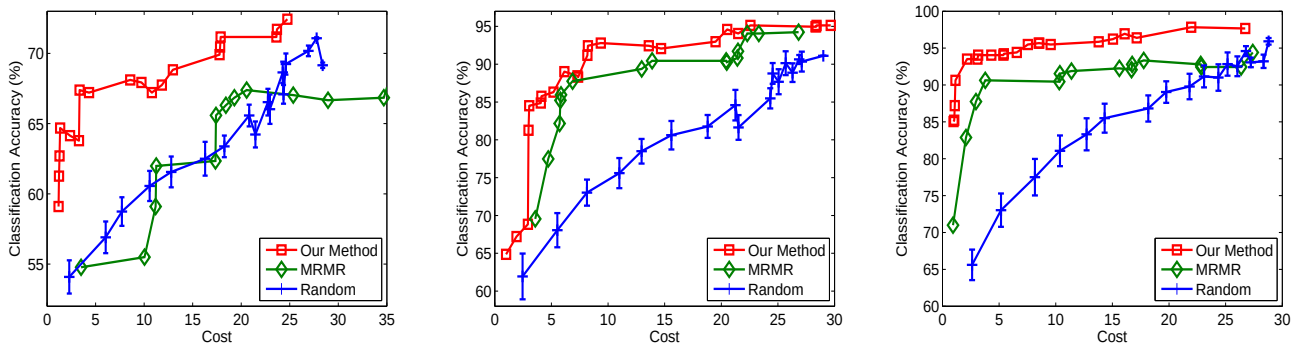

Fig. 3. The classification performance of different feature selection methods under different severities $\approx 10\%$ (left), $\approx 20\%$ (middle) and $\approx 30\%$ (right).

TABLE I
THE PHYSIOLOGICAL PARAMETERS USED FOR THE ARTERIAL TREE AND THEIR PHYSICAL MEANING

parameter	units	meaning	sample parameters for radial artery
L	cm	length of artery	23.5
R_p	cm	proximal radius of artery	0.174
R_d	cm	distal radius of artery	0.142
h	cm	arterial wall thickness	0.043
ρ	$g \cdot cm^{-3}$	blood density	1.05
μ	$g \cdot cm^{-1} \cdot s^{-1}$	blood viscosity	0.04
E	$10^6 \cdot g \cdot cm^{-1} \cdot s^2$	arterial wall elesticity	8

TABLE II
THE LIST OF ARTERIES REQUIRING LOW, MEDIAN, AND HIGH COSTS IN SENSOR DEPLOYMENT. NODE NUMBERS ARE THE SAME AS IN FIGURE 2.

Cost level	Cost Value	List of Arteries		
Low	0.03∼0.06	R. Carotid(5);	R. Radius(8);	R. Femoral(52);
		R. Subclavian B(7);	R. Post. Tibial(54);	L. Carotid(15);
		L. Radius(22);	L. Femoral(46);	L. SubclavianB(21);
		L. Post. Tibial(48)		
Median	0.8∼1.2	R. Subclavian A(4)	L. Subclavian A(19)	R. Ulnar A(9)
		L. Ulnar A(23)	R. Ulnar B(11)	L. Ulnar B(25)
		R. External Carotid(13)	L. External Carotid(17)	R. Interosseous(10)
		L. Interosseous(24)	R. Internal Carotid(12)	L. Internal Carotid(16)
		R. External Iliac(50)	L. External Iliac(44)	R. Common Iliac(42)
		L. Common Iliac(43)	R. Ant. Tibial(55)	L. Ant. Tibial(49)
		R. Deep Femoral(53)	L. Deep Femoral(47)	
High	4∼6	Ascending Aorta(1)	Aortic Arch A(2)	Aortic Arch B(14)
		ThoracicAorta A(18)	Thoracic Aorta B(27)	Brachiocephalic(3)
		Sup.Mesenteric(34)	Celiac A(29)	Celiac B(30)
		Abdominal A(28)	Abdominal B(35)	Abdominal C(37)
		Abdominal D(39)	Abdominal E(41)	L. Renal(36)
		R. Renal(38)	Gastric(32)	Splenic(33)
		Inf.Mesenteric(40)	Intercoastals(26)	L. Internal Iliac(45)
		R. Internal Iliac(51)	Hepatic(31)	R.Vertebral(6)
		Vertebral(20)		

[20] D. Tran and T. Nguyen, "Localization in wireless sensor networks based on support vector machines," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 19, no. 7, pp. 981 – 994, July 2008.

[21] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[22] F. Model, P. Adorjan, A. Olek, and C. Piepenbrock, "Feature selection for dna methylation based cancer classification," *Bioinformatics*, vol. 17, pp. 157–164, 2001.

[23] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomeld, and E. S. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, 1999.

[24] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1, pp. 273–324, 1997.

[25] B. Sch olkopf and A. J. Smola, *Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[27] I. Rodriguez-Lujan, R. Huerta, C. Elkan, and C. S. Cruz, "Quadratic programming feature selection," *Journal of Machine Learning Research*, vol. 11, pp. 1491–1516, April 2010.

[28] K. Zhang, I. Tsang, and J. T. Kwok, "Improved Nyström low-rank approximation and error analysis," in *International Conference on Machine Learning*, 2008, pp. 1232 – 1239.

[29] K. Zhang and J. T. Kwok, "Clustered Nyström method for large scale manifold learning and dimension reduction," *IEEE Transactions on Neural Networks*, vol. 21, pp. 1576–1587, 2010.

[30] J. Zhang and M. Small, "Complex network from pseudoperiodic time series: topology versus dynamics," *Physical Review Letters*, vol. 96, p. 238701, 2006.