Multiparameter physiological signal reconstruction using NARX neural networks

R. Matthew Wham, Xiaopeng Zhao

Abstract—Constant monitoring of a variety of physiological signals is vitally important in numerous clinical care settings. This signals are not perfect, however, and can be corrupted or lost. The loss of a signal can be devastating to the patient, as the physician may lose key information to understanding disease processes, or worse, be unaware of the patient’s status in either surgery or the ICU. This study uses a NARX-type Artificial Neural Network to reconstruct portions of physiological signals that have become corrupted. The effectiveness of this network was tested using signals and guidelines from the Computing in Cardiology/Physionet 2010 challenge, “Mind the Gap.” The NARX network performs quite well under these conditions, comparing favorably with other top entrants in the Physionet Cardiology/Physionet 2010 challenge, “Mind the Gap.” The effectiveness of this network also depends on which channel was corrupted. This work has important implications in many areas ranging from sports medicine and sleep studies to surgery and the ICU.

I. INTRODUCTION

Constant awareness of a variety of physiological signals is of paramount biological and clinical importance. These signals vary widely across medical specialties, but often include blood pressure, respiratory rate, electroencephalography (EEG), electromyography (EMG), plethysmography (PLETH), and electrocardiogram (ECG). At any one time, especially if a patient is in the Intensive Care Unit (ICU) or in surgery, knowledge, or lack thereof, of these physiological signals could mean the difference between life and death for the patient [1]. As a result, the loss or corruption of one of these signals could have a drastic impact on patient prognosis.

Many techniques have been developed for reconstructing signals of time series, including autoregressive moving average [2], nearest neighborhood prediction [3], adaptive filters [4], radial basis function [5, 6, 7], artificial neural networks [4, 8, 9], principal component analysis [10] and many others [11]. Here, we choose artificial neural networks (ANN) since they are robust and easy to implement, especially using the neural network toolbox in Matlab ® [12]. Neural networks, consisting of “neurons” (nodes) and “synapses” (connections between nodes), mimic the way the human brain operates [13]. ANNs are suited to problems of signal processing and reconstruction because they, like human brains, possess the ability to train, learn, and recognize patterns that conventional computer software might not. By changing the weights and biases they assign to each neuronal input, ANNs can “teach” themselves to minimize error and maximize useful output [13].

The primary tool for performance measurement of the ANN was guidelines set forth in a Physionet/Computing in Cardiology challenge known as “Mind the Gap” [14]. In this competition, teams of researchers were given a number of physiological records with data lasting ten minutes. Each record contains 6-8 channels ranging from ECG to a fingertip plethysmograph. In one of the channels, however, the last 30 seconds of signal was missing, resulting in a flatline. The object of the challenge was to use computer algorithms to reproduce the last 30 seconds of the missing signal, using information from the other intact signals. Two scores were used in Physionet to reflect the algorithm’s performance and accuracy of the reconstructions. To illustrate the definitions of the scores, we denote the target signals by time series {y_t} and the reconstructed signals by time series {ŷ_t}. One score, Q1, is calculated by normalizing the sum of the squares of residuals with the energy of the reference signal and then subtracting that from one: Q1 = max \left(1 - \frac{\text{MSE}(\hat{y})}{\sigma(y)^2}, 0\right), \quad (1)

where \(\sigma(y)\) represents the standard deviation of \(y\) and \(\text{MSE}(\hat{y})\) represents the mean square error. A good Q1 score indicates a good estimate of all point-to-point target signal levels. The second score, Q2, is calculated by finding the correlation coefficient between the reference signal and the reconstruction:

Q2 = \text{max}(\frac{\text{Cov}(y, \hat{y})}{\sigma(y)\sigma(\hat{y})}, 0), \quad (2)

where \(\text{Cov}(y, \hat{y})\) represents the covariance between the actual target and the predicted target. A good Q2 score indicates high correlation between the two [14].

The conference provided 3 sets of records, referred to as set A, B, and C. Set A was the only one out of the three for which the conference provided the target signal, thus, it was used in this study.

There are many different types of neural networks, including feed forward, feed forward backpropagation, cascade-forward backpropagation, radial basis function, and nonlinear autoregressive network with exogenous inputs (NARX). In addition, each of these networks has its own strengths, and thus ANNs have many wide-ranging applications. They have been used in many arenas such as attempting to classify cardiac events on ECGs, retail sales forecasting [15], and even predicting sky luminance in the tropics [16]. The plethora of applications for ANNs is a direct result of their ability to “learn” through training. Few other computing packages have the ability to adjust parameters and biases without human intervention in order to...
to optimize outcome and minimize error. This paper will focus specifically on applying the NARX network to signal processing.

II. METHODOLOGY
An artificial neural network (ANN), originally inspired by the structure and function of biological neural networks, is an adaptive computational model that can change its structure through learning. A neural network model consists of an interconnected group of artificial neurons. A neuron is a basic information-processing unit, which mimics the function of a biological neuron; see Figure 1. A neuron can accept multiple input signals through a set of synapses, which are modeled as weights. Positive weight values designate excitatory connections and negative values reflect inhibitory connections [4]. A summing junction adds up all the inputs according to their weights. An external bias is also added to the summing junction. Finally, an activation function controls the output of the neuron. In mathematical terms, the neuron model can be written as follows:

\[ y = \varphi \left( \sum_{i=1}^{N} w_i x_i + b \right) \]

Neural networks can adopt many different structures. In a network model, neurons are organized into layers. The input layer consists of source nodes, fed by delays of both the input and the output, which are projected onto an output layer of neurons through intervention of one or more hidden layers of neurons. Adding more hidden layers will enable the network to extract higher-order statistics, particularly when the size of the input layer is large. [17] Figure 2 shows a fully connected network since every node in each layer is connected to every other node in the next layer. If some of the communication links are missing, the network is said to be partially connected. A unique feature of the NARX schematic is the delay structure, creating embedded memory within the network, which is an important component of the NARX networks [18]; see Figure 2 for an example delay structure. NARX have a limited feedback [18]. In addition, NARX has two primary configurations: series-parallel and parallel. In its series-parallel structure, the input and the output are both fed back to the network. In the parallel configuration, however, only the input is fed to the network, and the output feeds back on itself. Shown below in Figure 2 is the series-parallel NARX structure, which is the type used in this study.

In order to improve the accuracy of prediction, we form a committee machine, where each expert in the committee is a NARX neural network. Figure 3 shows the block diagram of a committee machine. Here, a common input is used to train a number of different neural networks (experts). Then, outputs from individual experts are combined to produce an overall output through ensemble averaging. An alternative approach is to replace the committee machine with a single neural network. However, this will lead to a large number of adjustable parameters, significantly increase the training time, and may cause overfitting.

III. RESULTS
The primary tool for performance measurement of the ANN was guidelines set forth in a Physionet/Computing in Cardiology challenge known as “Mind the Gap” [14]. In this competition, teams of researchers were given a number of physiological records with data lasting ten minutes. Each record contains 6-8 channels ranging from ECG to a fingertip plethysmograph. In one of the channels, however, the last 30 seconds of signal was missing, resulting in a flatline. The object of the challenge was to use computer algorithms to reproduce the last 30 seconds of the missing signal, using information from the other intact signals. Two scores were used in Physionet to reflect the algorithm’s performance and accuracy of the reconstructions. To illustrate the definitions of the scores, we denote the target signals by time series \( \{ y_t \} \) and the reconstructed signals by time series \( \{ \hat{y}_t \} \). One score, Q1, is calculated by normalizing the sum of the squares of residuals with the energy of the reference signal and then subtracting that from one: 

\[ Q1 = \max \left( 1 - \frac{\text{MSE}(\hat{y})}{\sigma(y)^2}, 0 \right), \]

where \( \sigma(y) \) represents the standard deviation of \( y \) and \( \text{MSE}(\hat{y}) \) represents the mean square error. A good Q1 score indicates a good estimate of all point-to-point target signal levels. The second score, Q2, is calculated by finding the
correlation coefficient between the reference signal and the reconstruction:

$$Q_2 = \max \left( \frac{\text{Cov}(y, \hat{y})}{\text{Var}(y)}, 0 \right),$$  \hspace{1cm} (2)

where Cov(y, \hat{y}) represents the covariance between the actual target and the predicted target. A good Q2 score indicates high correlation between the two [14]. The conference provided 3 sets of records, referred to as set A, B, and C. Set A was the only one out of the three for which the conference provided the target signal, thus, it was used in this study.

![Figure 4](image)

Figure 4. Aggregate scores for all 100 records in set A: straightforward simulation (triangles) and accumulated averaging (circles).

We compute the aggregate scores for the 100 records; see Figure 2. Here, results marked with triangles (blue) are obtained using straightforward simulations of the NARX network. Note that we repeat the computation for 10 trials. In most trials, the aggregate Q1 scores are within the range of 52 to 64 whereas the aggregate Q2 scores are in the range of 70 to 78. Results marked with circles (red) are obtained using the committee machine method. Here, we calculate the average of predictions of the first n iterations. Thus, the average accumulates when n increases. Moreover, aggregate Q1 and Q2 scores of the accumulated averaging mechanism smoothly converge as the number of iterations increases. The converged Q1 and Q2 scores are improved by more than 6 points, compared to results of straightforward simulations. We carry out extensive simulations using various parameters and consistently observe that iterative retraining and accumulated averaging significantly improve the accuracy of the prediction. The trends are similar to those shown in Figure 4. We note that the results here are comparable to some of the best results presented in the Physionet 2010 Challenge, especially in Q2 scores.

![Figure 5](image)

Figure 5 shows the time histories of the actual targets and the predicted targets of a few selected records. All reconstructions in this figure are obtained from accumulated averaging of 10 iterations. The top panel shows a CVP target. Here, the recordings in the actual target are flatted out by the upper limit of machine reading. Although the network model is not able to recover the flat saturation in the signal, it faithfully captures the overall trend, producing scores of Q1=0.56 and Q2=0.80. The bottom panel shows a time history of an ECG target, which shows a transition from normal rhythms to abnormal rhythms. The second half of the record clearly shows saturation in reading and chaotic patterns of arrhythmias. The model produces a reasonable reconstruction with Q1=0.53 and Q2=0.83. It is interesting to note that the neural network models can accurately reconstruct physiological records even when the signals are highly random and irregular.

### IV. CONCLUSION

We have explored the application of NARX networks in predicting physiological signals using data records from Physionet 2010 Challenge. Since a NARX network is a highly nonlinear model, the training process, starting from randomly generated initial guesses, usually converges to local optimal solutions. As a result, different trials are likely to produce different predictions. Indeed, scores from different trials range between 52 and 64 in Q1 and between 69 and 79 in Q2.

While the direct results from NARX networks are of decent accuracy, the prediction accuracy can be significantly improved by forming a committee machine, which produces final predictions using average of the results from several different trials. When the number of experts in the committee machine increases, the scores of the committee increase as well. Finally, the scores saturate at about 71 for Q1 and 86 for Q2. The NARX based committee machine is able to produce accurate predictions for BP, CVP, and ECG signals whereas the predictions for PLETH and RESP signals are not as well. Particularly, the Q1 scores of RESP signals average at 50. This is probably because the RESP

<table>
<thead>
<tr>
<th>Target</th>
<th>Q1</th>
<th>Q2</th>
<th>Num. of records</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>0.865 +/- 0.129</td>
<td>0.943 +/- 0.067</td>
<td>10</td>
</tr>
<tr>
<td>CVP</td>
<td>0.658 +/- 0.272</td>
<td>0.834 +/- 0.145</td>
<td>10</td>
</tr>
<tr>
<td>ECG</td>
<td>0.832 +/- 0.219</td>
<td>0.932 +/- 0.139</td>
<td>44</td>
</tr>
<tr>
<td>PLETH</td>
<td>0.615 +/- 0.276</td>
<td>0.769 +/- 0.251</td>
<td>15</td>
</tr>
<tr>
<td>RESP</td>
<td>0.494 +/- 0.292</td>
<td>0.745 +/- 0.179</td>
<td>21</td>
</tr>
<tr>
<td>All</td>
<td>0.714 +/- 0.283</td>
<td>0.860 +/- 0.185</td>
<td>100</td>
</tr>
</tbody>
</table>
signals vary much more slowly than other signals, which are used to predict the RESP signals. Thus, we hypothesize that the prediction of RESP signals can be improved by low-pass filtering the input channels before applying the network model and the committee machine. We further note that the developed model is robust to noise, chaos, and irregularity in the signals. To apply this technique for medical diagnosis, the results must be carefully reviewed by clinicians to decide what Q1 and Q2 values are acceptable. Nevertheless, based on our observations of various numerical results, the predictions are visually very close to the targets when Q1 is about 0.7 or Q2 is 0.8.

Figure 5. Selected examples of target signals and reconstructions: CVP (top left, Q1=0.56 and Q2=0.80), ECG (top right, Q1=0.53, Q2=0.83), PLETH (bottom left, Q1=0.87, Q2=0.93), and respiration (bottom right, Q1=0.64, Q2=0.80).

ACKNOWLEDGMENT

The authors are grateful to Joseph McBride, Adam Sullivan, and Henian Xia for their insightful discussions. The authors thank Gerald Ragghianti for technical support on cluster computing.

REFERENCES