Neural Networks for ECG Compression and Classification

I Habboush, GB Moody, RG Mark

Harvard-MIT,
Division of Health Sciences and Technology, Cambridge, MA

Abstract

The neural network model provides a framework for studying applications of nonlinear transformations to signal processing. We compared neural networks designed for ECG compression and classification with optimum linear methods. We found that simple neural networks with one hidden layer approach the performance of linear methods but offer no advantage over them. Suitably constructed networks with more than one hidden layer, however, can perform more efficient ECG compression than is possible using linear methods under the same constraints.

1 Introduction

Artificial neural networks (ANNs) are of interest in signal processing applications because they can be trained to extract the essential features from a set of known signals, and then can use these features to process unknown signals. Their major strengths lie in their ability to recognize or correctly classify patterns which have never been presented to the network before. Much of the recent interest in neural computing has resulted from the development of effective algorithms for designing and training networks of various topologies [1] [2].

In signal processing, neural networks have proven useful in classification and dimensionality reduction. Iwata *et al.* [3] have described a neural network which can compress an ECG which it has been trained to represent. Typically, an ANN constructed for this purpose has many input units, an equal number of output units, and a small number of hidden units [4]. It is trained to perform the identity transform, i.e., the network parameters are iteratively adjusted until the outputs closely approximate the inputs for all of the signals in the training set. Once this has been done, the network parameters are recorded, the input and hidden layers are used for compression, and the output layer is used for reconstruction. To compress the signals, they are presented to the input layer, and the outputs of the hidden layer are recorded. At any later time, the recorded hidden-layer outputs may be presented to the output layer to obtain a reconstruction of the original signals. The compression achieved is a function of the ratio between the number of inputs and the number of hidden units (neglecting the overhead needed to record the network parameters).

Iwata *et al.* presented each cardiac cycle as a set of inputs to their network, which was able to reconstruct these inputs using the recorded outputs of only two hidden units. A limitation of this method is that the network must be trained anew for each subject's ECG, and for new morphologies within each subject's ECG. In consequence, there must be an alternate means of storing cardiac cycles which are poorly represented by the network; if the number of such cycles is relatively small, they may be stored in uncompressed form without significant penalty. Another significant limitation is the requirement that the network be trained "on the job" (i.e., while it is being used for compression), which implies the need for much more computation than is needed to run a fixed network, as well as a modest increase in storage requirements in order to maintain information about changes in the network parameters.

Iwata's network may be considered as a nonlinear transformation of the input into a pair of basis functions (represented by the hidden units) which are then recombined in a nonlinear fashion by the output layer into an approximation of the input. Just as in the case of linear transformations such as the Karhunen-Loève transform (KLT), one may expect that a wider variety of inputs should be representable by increasing the number of basis functions. We have explored this idea and tested the feasibility of building a multi-layer system with fixed parameters which can efficiently compress an ECG which had not been used to train it.

Since neural networks are capable of general nonlinear transformations [5], such a study offers the oppor-

0-8186-2485-X/92 $3.00 © 1992 IEEE

tunity to determine if a nonlinear transform has any advantage in this application over the truncated KLT, which is the optimum linear method for data reduction [6]. Although no optimum nonlinear transformation can be identified *a priori*, we can begin with the observation that a neural network can be constructed to implement the truncated KLT, so that the data reduction achievable using the truncated KLT is a lower bound on the best results obtainable using a neural network. In our study, we compared the performance of our ANNs with that of the truncated KLT, using in each case the same training set and equal numbers of hidden units and KL basis vectors in order to obtain a meaningful comparison.

It should be noted that we did not attempt to address the question of compressing the entire ECG; rather, we focused (as did Iwata *et al.*) on compressing the P-QRS-T only. For a practical application of these techniques, it is necessary to have a reliable QRS detector for identification of the waveforms to be compressed, and to use another method to represent the segments which fall between the detected P-QRS-T waveforms. It is critically important to do so, since isolated P-waves (as well as entire cardiac cycles missed by the QRS detector) must be represented in the output. In most cases, a piecewise linear representation is a very efficient solution to this problem [7].

2 Methods

A *training set* of 78 beats was selected from the MIT-BIH Arrhythmia Database. The training set included a wide range of morphologies. A fiducial point, FP, was placed near the R-wave peak of each beat. One hundred samples were used to represent each beat: 15 samples during the P-wave (from $FP-260$ ms to $FP-175$ ms), 60 samples during the PQ segment, the QRS complex, and the ST segment (from $FP-175$ ms to $FP+175$ ms), and 25 samples during the T-wave (from $FP+175$ ms to $FP+330$ ms). We sampled the QRS complex at a higher frequency than either the P-wave or the T-wave in order to preserve high-frequency detail there.

Two other data sets were also prepared for development and evaluation. The *development set* of 165 beats, also from the MIT-BIH Database but disjoint from the training set, was used throughout the study to assess the performance of a wide variety of ANNs and to guide their development. The *evaluation set* of 271 beats (130 from the AHA Database, and 141 from the European ST-T Database) was used at the end of the study to test the performance of the ANNs which appeared to be the most successful.

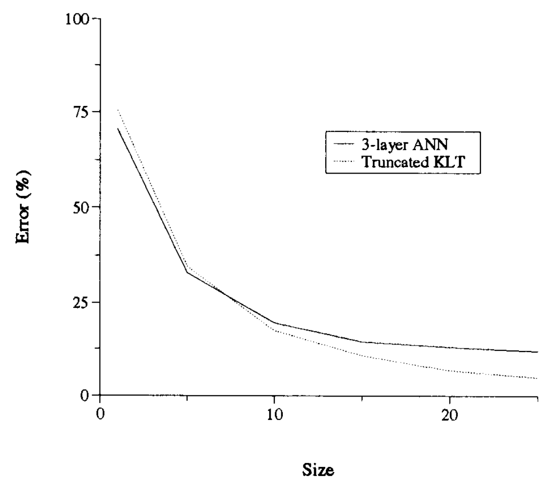

Figure 1: Errors decrease for neural networks as the number of hidden units increases, and for the truncated KLT as the number of coefficients increases.

We used a commercial neural network simulator [8] to implement our neural network architecture. A three-layer backpropagation algorithm was used in most simulations [9]. The input and output layers were fixed to have 100 units, corresponding to the 100 samples per cardiac cycle as described above. The size of the hidden layer varied for different simulations.

3 Results

The reconstruction error is a monotonically decreasing function of the number of KLT basis vectors and the size of the hidden layer of the neural network. Although a perfect reproduction of the input can be achieved using all of the KLT bases (with no net data reduction), obtaining a similar result from a neural network with a large hidden layer may require an impractically large number of training cycles and may depend on other factors [10]. The networks in figure 1 were trained for 5000 cycles. While an M-hidden unit network performs better than a truncated KLT with M coefficients for small M, the reverse is true for large M. Since the length of training needed may be expected to increase with M, insufficient training may explain this observation.

To test this hypothesis, we studied a network with $M = 15$ in detail. This network has a hidden layer large enough to produce a good approximation of its input (see figure 1), while being small enough to train easily

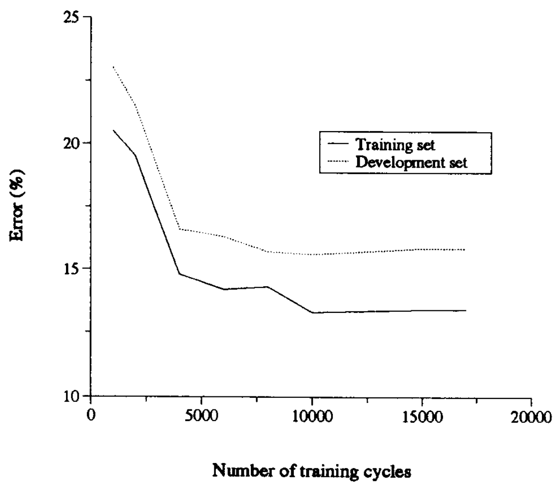

Figure 2: Errors for a three-layer ANN with 15 hidden units as the number of training cycles varies.

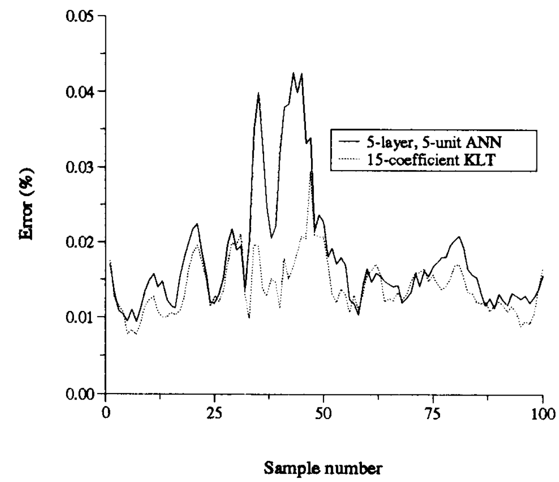

Figure 3: Errors for a five-layer ANN with 5 hidden units in the middle layer.

	Development set			Evaluation set		
	P	QRS	T	P	QRS	T
ANN	66.8%	14.0%	20.7%	70.3%	18.0%	31.0%
KLT	58.7%	11.2%	16.1%	55.6%	12.2%	21.4%

Table 1: Errors in beat reconstruction for $M = 15$ (see text).

and to permit reasonable compression, as compared with larger values of M.

Figure 2 shows the error function of a 15 hidden unit network as the number of training cycles varies. Little change occurs after roughly 10000 cycles. Examination of the errors obtained in reconstructing the development set shows monotonically decreasing errors without evidence of overtraining [11]. At this stage, we expect the neural network to give optimum performance for this configuration.

Table 1 shows the results of reconstructing beats from both the development (MIT-BIH) and the evaluation (AHA and EDB) sets. Root mean square errors (RMSEs) were calculated for the P-wave, QRS complex, and T-wave of each beat (normalized by the areas of the respective segments). From such calculations, we can compare the error in reconstructing the P-wave, for instance, to the area of the P-wave in a single beat. Similar comparisons are made for the QRS complex and the T-wave. The relatively small error percentages for the QRS complex in the table reflect the size of the errors relative to the size of the QRS complex, and not in any absolute sense. The truncated KLT shows a slight advantage over the neural network for $M = 15$. Based on the results shown in figure 2, there is no reason to believe that this advantage would be diminished if the network were trained longer.

Thus, we found that the 3-layer ANN did not have any advantage over the Karhunen-Loève transform. When we investigated ANNs with larger numbers of layers, however, we found some which do appear to perform more efficient data reduction than is possible using the KLT. Figure 3, for example, shows the performance of a 5-layer ANN with only 5 hidden units in the middle layer. The figure also shows the performance of a 15-coefficient KLT reconstruction of the AHA-EDB data set. In the figure, the RMSE for every sample of each beat in the data set is normalized by the area of that beat. Although the error during the QRS complex is roughly 2% higher for the 5-layer ANN than for the KLT, errors during the P- and T-waves are not significantly different, even though $M = 5$ for the ANN and $M = 15$ for the KLT.

For a trained neural network, the activations of the hidden units are characteristics of each beat. Thus we may expect these values to provide a way to classify morphologies. Figure 4 illustrates how a 3-layer network with $M = 2$ might be used for morphologic analysis, using 1000 beats from AHA Database record 5002. For comparison, the first two KL coefficients are plotted in the same way. Similar clustering is visible in both plots. Preliminary results from our study of this problem suggest that clustering using hidden-layer

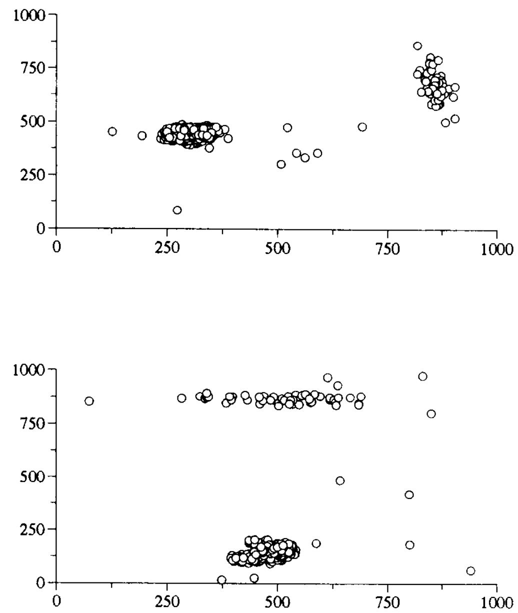

Figure 4: Clustering using 2-unit ANN (top) and 2 KL coefficients (bottom).

outputs of ANNs may work as well as clustering using the KLT, with the advantage that more efficient compression can be performed at the same time, without significantly different computational effort.

4 Conclusions

By this study, we hoped to answer a few basic questions about the nature of data reduction transformations using neural networks. Although a purely experimental approach is not sufficient to derive general properties of ANNs, we have shown that neural networks are capable in some circumstances of more efficient ECG data reduction than is possible using linear methods operating under the same constraints. Although we found that fixed 3-layer networks offer no advantage over the KLT, our results using networks with more than 3 layers indicate the opposite. The difficulty in studying these networks, as well as in applying them, is substantial, however: large computational demands and very low convergence rates are characteristic of many-layered networks. Current research activities in the area of neural computing are concentrated on developing more efficient algorithms and VLSI neural network implementations, which may make it possible to explore very large and complex neural architectures.

References

[1] Lippman, R.P. An introduction to computing with neural nets. *IEEE ASSP* pp. 4–22 (April 1987).

[2] Hecht-Nielsen, R. Theory of backpropagation neural networks. *International Joint Conference on Neural Networks*, : 593–608. San Diego: SOS Press (1989).

[3] Iwata, A., Nagasaka, Y., and Suzumura, N. Data compression of the ECG using neural networks for digital Holter monitor. *IEEE Trans. Eng. in Med. and Biol.* pp. 53–57 (September 1990).

[4] Cottrel, G.W., Munro, P., and Zipser, D. Image compression by back propagation: an example of extensional programming. In *Advances in Cognitive Science*, Vol.3, ed. N.E. Sharkey. Norwood, NJ: Ablex (1987).

[5] Kreinovich, V.Ya. Arbitrary nonlinearity is sufficient to represent all functions by neural networks: a theorem. *Neural Networks* **4**: 381–383 (1991).

[6] Moody, G.B., and Mark, R.G. QRS morphology representation and noise estimation using the Karhunen-Loève transform. *Computers in Cardiology* **16**:269–272 (1989).

[7] Moody, G.B., Soroushian, K., and Mark, R.G. ECG data compression for tapeless ambulatory monitors. *Computers in Cardiology* **14**:467–470 (1987).

[8] *ANSIM*. San Diego, CA: Science Applications International Corp. (1987).

[9] Rumelhart, D.E., and McClelland, J.L. *Parallel Distributed Processing*, Volume 1. Cambridge, MA: MIT Press (1986).

[10] Hornik, K. Multilayer feedforward networks are universal approximators. *Neural Networks* **2**:359–366 (1989).

[11] Hecht-Nielsen, R. *Neurocomputing*. Reading, MA: Addison-Wesley (1990).