Data Compression of ECG Signals using Error Back Propagation (EBP) Algorithm

Anuradha Pathak, A. K. Wadhwani

Abstract: ***Heart is one of the vital parts of our human body, which maintains life line. The paper deals with an efficient composite method which has been developed for data compression and signal reconstruct of ECG signals. ECG data compression algorithm is needed that will reduce the amount of data to be transmitted, stored and analyzed, but without losing the clinical information content. After carrying out detailed studies and by training different topologies of error back propagation (EBP) artificial neural network (ANN) with respect to variation in number of hidden layers and number of elements, the topology with single hidden layer and four elements in each hidden layer has been finalized for ECG data compression using a Physionet.org data base. The compression ratio (CR) in ANN method increases with increase in number of ECG cycles. The entire programming in this paper is carried out on the version of MATLAB 7.8.***

Index Items- Compression, Data compression, ECG, Compression ratio (CR), PRD and EBP.

I. INTRODUCTION

ECG is a very important physiological parameter associated with cardiac disorders; it is not necessarily used only at hospitals but also at distant places. In the cases like critical cardiac patients, ambulatory patients and people other cardiac surveillance, it is not possible to transmit the entire ECG data; the ECG signal is recorded and transmitted to a distant location continuously, so compression of ECG data becomes necessary. The continuing proliferation of computerized electrocardiogram (ECG) processing systems along with the increased feature performance requirements and demand for lower cost medical care have mandated reliable, accurate, more efficient ECG data compression techniques. The practical importance of ECG data compression has become evident in many aspects of Computerized electrocardiography including: a) increased storage capacity of ECG's as databases for subsequent comparison or evaluation , b)Feasibility of transmitting real-time ECG's over the public phone network, c)Implementation of cost effective real-time rhythm algorithms, d)Economical rapid transmission of off-line ECG's over public lines to a remote interpretation center, and e) Improved functionality of ambulatory ECG monitors and recorders. The ECG data is compressed with a view to storing it and later on decompressing it without any loss of diagnostic

information. Compression is a way to reduce the no. of bits in a frame but retaining its meaning. Decreases space, time to transmit and cost. Technique is to identify redundancy and eliminate it. The idea of represent is signal/information in fewer bits and any signal that contains some redundancy to be compressed. A compressor can reduce the size of a file by deciding which data is more frequent and assigning it less bits than to less frequent data to save time when transmitting it and to save space when storing it . The design of data compression scheme therefore involves trade-offs among various factors including the degree of compression, the amount of distortion introduced (if using a lossy compression scheme) and the computational resources required to compress and uncompress the data . Data compression methods are lossless (text or program) & lossy (image, video, audio). Lossless compression: In lossless method, original data and the data after compression and decompression are exactly same. Lossless methods are used when we can't afford to loss any data; legal and medical documents, computer program. Ex- WIN ZIP program. Lossy compression: Loss of information is acceptable in a picture of videos .Loss of information is that our eyes and ears can distinguish stable changes. Loss of information is not acceptable in a text file or program file. Ex- JPEG, MPEG. Data compression of ECG signal has been carried out using error back propagation to use (EBP)artificial neural network (ANN).The parameter has been performed both the original signal(before compression), and on the reconstructed signal (after decompression). Fig. 1 shows the complete details of the overall process of data compression. The data created from physionet.org database. The sampling frequency of the physionet database is (100 samples per ECG cycles). The ECG data (training data) with possible variations in the signals were used to train the ANN and fix the values of weights of hidden layers and output layers. Once the training had been Completed, the trained network was tested for its data compression performance by test data of the ECG signal.

Manuscript received March 31, 2012

Anuradha Pathak , Student M.E (2nd year), Measurement and Control systems, Electrical Engineering Department, MITS , Gwalior, Madhya Pradesh, India. (E-mail-anuradha.pathak27@gmail.com).

Dr. A.K Wadhwani, Professor Electrical Engineering Department, MITS, Gwalior, Madhya Pradesh, India,
(E-.mail:wadhwani_arun@rediffmail.com).

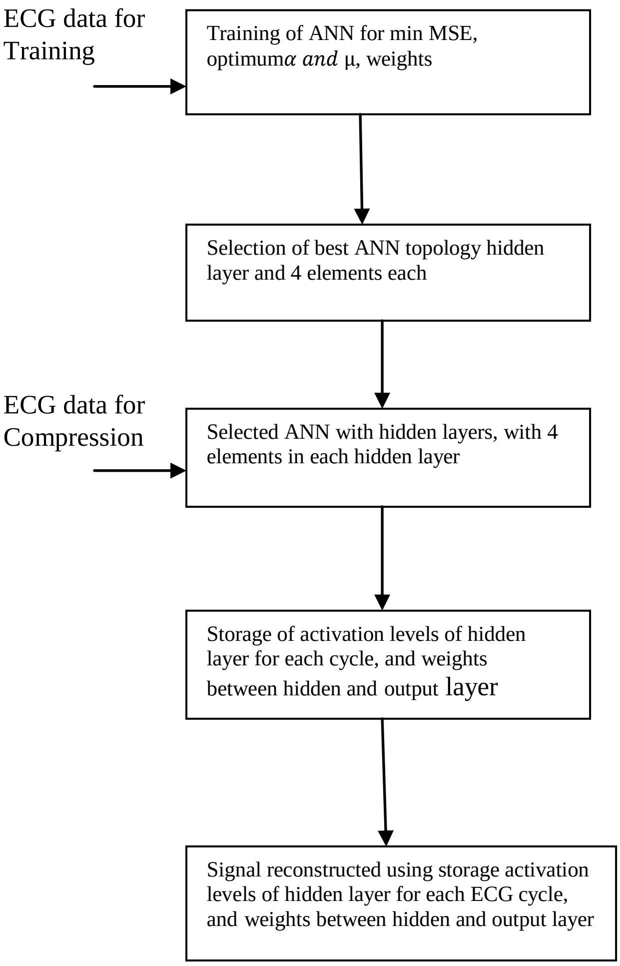

Figure 1. General flow-chart of Composite method

1.1 ECG Compression techniques

Many compression algorithms exist have shown some success in electrocardiogram compression; however, algorithms that produce better compression ratios and less loss of data in the reconstructed data are needed. It will provide an overview of several compression techniques and algorithms that should improve compression ratios and less error in the reconstructed data. Finally, an easy to use computer program will be written

in "MATRIX LABORATORY (MATLAB)", which will allow its user to compare various compression schemes and analyze reconstructed electrocardiogram records through a graphic interface, without detailed knowledge of the Mathematics behind the compression algorithms and find the various waveforms.

II. DATA COMPRESSION

Data compression techniques have been classified in a broad spectrum of communication areas such as speech, image and telemetry transmission. Biomedical signals can be broadly classified into three categories: direct data compression and transformation methods. Direct data compression techniques are ECG differential pulse code modulation and entropy coding, turning point (TP), amplitude-zone-time epoch coding (AZTEC), coordinate reduction time encoding system (CORTES), Fan and SAPA algorithms. Some of the Transformational approaches are discrete cosines transformation (DCT), fast Fourier transforms (FFT), and discrete sine transforms (DST).

2.1 Direct ECG Data Compression

There are a number of direct data compression techniques (Satch et. al.1990) the amplitude zone time epoch coding (AZTEC) algorithm is a popular data reduction technique with an achieved compression ratio of 10:1 for an ECG signal sampled at500 Hz with 12 bit resolution (Cox et al.1972,1974).However, the reconstructed signal demonstrates significant discontinuities and distortion occurs in the reconstruction of P and T waves due to their slowly varying nature(Satch et al. 1990,Furht and Perez 1988).Later on a modified AZTEC algorithm with adaptive error threshold was proposed which resulted in a better reconstructed signal. AZTEC algorithm adaptive error threshold was proposed which resulted in a better reconstructed signal. The TP algorithm was reducing the sampling frequency of the ECG signal from 200 to 100 Hz without diminishing the evaluation of large amplitude. The compression ratio of this method is 2:1 and the reconstructed signal resembles the original signal with small distortion .The CORTES (coordinate reduction time epoch coding) is a combination of ATEC and TP methods and applies a parabolic smoothing to Aztec Portion of the reconstructed signal for reducing the distortion .The Fan method was originally reported and tested on ECG signals by Gardenhire (1964,1965).Gardenhire compared the FAN performance with that of the step method, concluding that the FAN method provides the best performance in respect of both the compression and the signal fidelity.Ishijima et al. (1998) presented three scan along polynomial approximation (SAPA) algorithms, namely SAPA-1,SAPA-2 and SAPA-3.SAPA-1 converges upper and lower bound slopes within permissible threshold limits for successive data points. When any succeeding point falls outside the converging slopes, the new data points is saved as a reference points and the process is continued . SAPA-2 makes use of the center point slope along with the upper and lower bound slopes, SAPA-2 producer better results for ECG signals (Satch et al. (1990). SAPA-3 which is a combination of SAPA-1 and SAPA-2 ,has the greatest compression efficiency and longest execution time, Where as SAPA-1 has the least compression efficiency and shortest execution time (Ishijima et al. 1990) . The DPCM (delta pulse code modulation) techniques provides a compression ratio of 10:1 for an ECG signal sampled at 100 Hz .In an improved implementation, DPCM with a 300 Hz Sampling rate provided a compression ratio of 4:1 (Stewart et al. 1973), The peak picking scheme of direct data compression is more suitable for fast changing segments of ECG signals (Imai et al. 1985), Ibiyemi 1986, Lachiver et al. 1986, Giakoumakis and Papkonstantinou 1986). The cycle to cycle compression method, which is basically a template matching technique, shows no improvement over the FAN compression method for ECG signal (Satch et al. 1990) Many transform techniques have been used for compression of ECG signals (Satch et al. 1990).FFT (Fast Fourier transform), KL transform (KLT), Walsh transform (WT), Harr transform (HT) and cosine transform (CT) have been successfully used in data compression.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Data modeling and reduction are also performed by a number of methods like Principal component analysis (PCA), Multi variate adaptive regression (MARS), Adaptive resonance theory (ART), Kohonen's self organizing map (KSOM) and error back propagation artificial neural network (EBP-ANN).

Learning can be done in supervised or unsupervised manner. In supervised learning, both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting output against the desired outputs. Errors are then calculated, causing the system to adjust the weights which control the network. In unsupervised learning, the network is provided with its inputs but not with desired outputs. The system itself must then decide what features it will use to group the input data .this is often referred to as self organization or adaption. The regression methods have been compared in one way or another with the sigmoidal back-propagation neural network, regarding their ability to classify, reduce and reconstruct original data, and EBP-ANN has been largely recommended for the compression of data like an ECG signal, error back propagation sigmoidal ANN technique has been used for ECG data compression. The number of input layer neurons of the selected ANN being equal to the number of samples of ECG in one cycle, all cycles having the same no. of samples, depending on the sampling frequency.

III. EBP-ANN BASED DATA COMPRESSION

ANN's are used successfully for data compression of signals of several different types .For this purpose the network is first trained for the signal to be compressed. A large amount of signal is presented to the network on a cycle to cycle basis .After the completion of successful training the trained weights between the output and hidden layer and the activation levels of the hidden layer units are stored during compression. The stored information requires less space and is used for reconstruct of the original signal. The ANN method of data compression does not reduce the number of samples of the original signals, but retains some key parameters, i.e. activation levels of hidden layers and connection weights, to represent the original signal is compressed form. Of the activation level of hidden layers and connection weights to represent the original signal is compressed form. Of the existing ANN's the error back propagation (EBP) ANN is the most topology for ECG data compression. The EBP-ANN with single hidden layer structure is shown in Fig.2. ECG data compression is accomplished after training .The details of the algorithm for training and data compression are as follows.

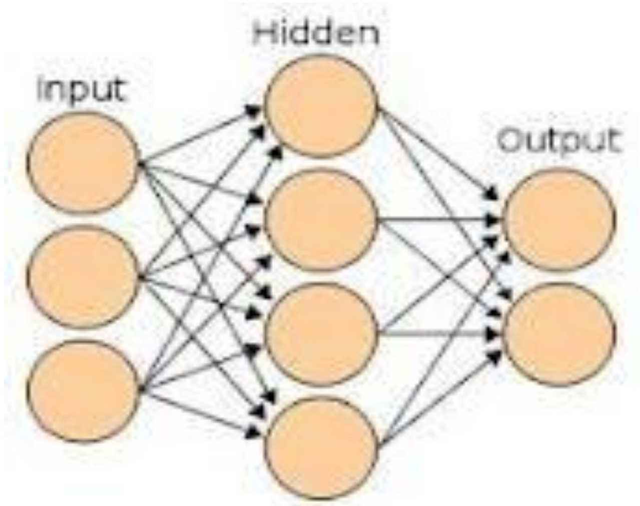

Figure 2. Single layer ANN structure

3.1Training Steps

The following steps are taken during the training phase.

(a) Cycles of the ECG, from the data to be compressed, are taken as training patterns.

(b) These training patterns are presented to the network on a cycle-to-cycle basis. During presentation of this signal the activation levels of the nodes in each layer are calculated.Net

input to any node j, connected to input nodes of the previous layer is given by the expression

$$i_j = \sum_{i=o}^{p} W_{ij} O_i$$

Where W_{ij} is the weight connecting node i to node j and p is the number of input nodes. The output or activation level node j is the sigmoidal function of its net input

$$O_j = 1/\ [1+\exp(-i_j)]$$

Similar equations can be written for the net inputs and output / activation levels of nodes in other layers of this feed forward network.

(c) Simultaneously, adjustment of weights between all layers of the network is carried out using the equation.

$$W_{ij}(new) = W_{ij}\ (\mathrm{Old}) + \mu \delta_j O_i + \alpha\ [\Delta W_{ij}\ (\mathrm{old})]$$

Where δ_j is the error term for the jth node in the hidden or output layer, propagated back to perform weight adjustment. The coefficient μ and α represent gain coefficient and momentum coefficient.

(d) The above steps are repeated for each training cycle in the training pattern, and after every presentation of the whole training pattern the mean square error (MSE) is calculated.

$$MSE=\frac{1}{N}\sum_{(t_n-o_n)} 2$$

Where t_n is the target value, o_n is the output value for the nth output node and N is the number of training cycles.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

(e) The MSE is supervised during each presentation or iteration of the trained pattern. The network is trained for repeated presentation until the reduction is MSE almost ceases.

3.2 Compression

After completing the training, the input data for compression is presented to the trained network on a cycle-to-cycle basis. For each cycle of input data the activation levels of different layers are calculated, the difference is that the weights are taken as the trained weight are stored after training. The algorithm for training was tested on the physionet data base for different topologies of the network with respect to variations in the hidden layer.ANN with single hidden layer having two and four elements was trained. It was found that a network with single hidden layers, having four elements in each layer, a gain coefficient of 0.75 and momentum coefficients of 0.85, produced the best results for ECG data sampled at 100 Hz.

Table 1. Training results with single hidden layer

No. of hidden units	Gain coefficient	Momentum coefficient	Total MSE at termination	No. of iterations
5	0.0045	0.60	0.0771	6236
10	0.035	0.75	0.1455	2696

IV. SIGNAL RECONSTRUCTION

For signal reconstruction, the outputs of the output layer nodes are calculated as

O_n =1/ [1+exp. (-i_n)]

Where in $=\sum_{j=0}^{q} W_{jn} O_j$

Here W_{jn} and O_j are the values stored during compression and q is the number of nodes in hidden layer.

The above step is repeated for each set of the stored values corresponding to different cycle of the input data. The output values thus obtained are the samples of reconstructed signal.

V. PERFORMANCE EVALUATION

The convergence of error and the comparison of the morphologies of the original and reconstructed signal.MSE is thus the main performance index for check the quality of data compression; the main factor is that the lower the MSE, the better is the training and hence the closer is reconstructed signal to original signal. The ANN not only performs the function of data compression but also eliminates the present noise from the original signal. Here, the present original and reconstructed signal is same so PRD value has not calculated.

5.1 Compression ratio

During compression, activation levels of the hidden layer units are stored for each cycle of the input data and the weights between the hidden layer and output layer are to be stored only once for all cycle .The compression ratio depends upon the number of ECG cycles compressed The more cycle compressed ,the compression ratio is large. For n=100, i.e. 100 cycles of ECG signal. The ANN method, if proper trained the network then the signal reconstruction is better than the other methods.

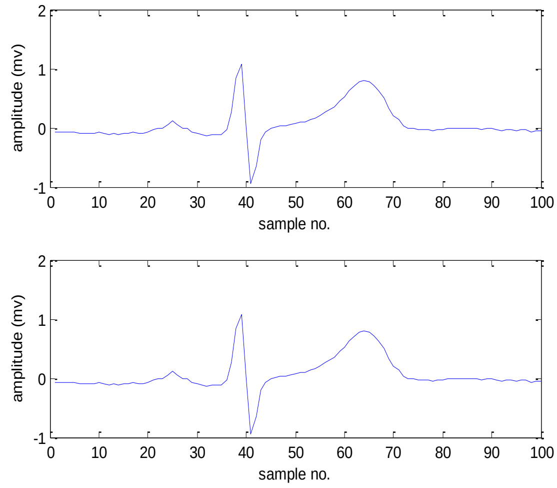

FIGURE 3. SINGLE ECG WAVE SEGMENT
(a) Original signal
(b) Reconstructed signal

Table 2.Comparison of different data compression techniques for 100 ECG cycles

Method	Compression ratio	Sampling frequency(Hz)	PRD (%)
AZTEC	10.0	500	28.0
TP	2.0	200	5.3
CORTES	4.8	200	.0
FAN/SAPA	3.0	250	4.0
Entropy coding of second difference ECG	2.8	250	-
Peak-picking (spline) with Entropy coding	10.0	500	14.0
DPCM-Delta with threshold	4.0	300	-
DPCM-linear prediction	2.5	500	-
Orthogonal transform (CT,KLT,HT)	3.0	250	-
Dual application of K-L transform	12.0	250	-
Fourier description	7.4	250	7.0
Proposed method	20	360	-

VI. RESULT & DISCUSSION

In this paper, we have presented an effective ECG Data Compression algorithm to be implemented in MATLAB based on a newly developed tool i.e. error back propagation Artificial Neural Network (EBP-ANN) to the processing of ECG and various bioelectrical signals. Bioelectrical signals like ECG .which can be change into equal samples of cycles, can be compressed and reconstructed successfully. The compression ratio (CR) in ANN method increases with increases in number of ECG cycles. In future we applying the same method in this topic but we find the Features of ECG signals in original signals and reconstructed signals.

ACKNOWLEDGEMENT

The authors are thankful to the UGC for their financial support under major research project sanctioned by UGC New-Delhi.

REFERENCES

1. S.M.S. Jalaleddine., "ECG Data Compression Techniques - A Unified Approach", IEEE Trans. on Biomed. Eng., Vol. 37, No. 4. 1990, pp. 329-343.
2. Y. Zigel, A. Cohen, and A. Katz, "ECG Signal Compression Using Analysis by Synthesis Coding", IEEE Trans. on Biomed. Eng, Vol. 47, No. 10, October, 2000, pp. 1308-1316;
3. Y.Kocyigit, "ECG Data Compresion by Artifical Neural Networks," Master Thesis, I.T.U., 1996.
4. S.C. Saxena, Vinod Kumar and V.K.Giri, 'ECG Data Compression using EBP-NN "Vol. 20,No.6, Nov-Dec 2003,pp 583-604.
5. S C Saxena, A Sharma & S C Chaudhary, Data compression and feature extraction of ECG signals, International Journal of System Science, Vol. 28, No. 5, pp 483-498, 1997.
6. J R Cox, F M Nolle , H A Fozzard & G C Oliver, AZTEC A preprocessing program of real time ECG rhythm analysis, IEEE Trans on BME vol. 15, pp 128-129, 1968.