

DATA COMPRESSION USING NEURAL NETWORKS IN BIO-MEDICAL SIGNAL PROCESSING

Mandavi¹, Prasannjit², Nilotpal Mrinal³, Kalyan Chatterjee⁴ and S. Dasgupta⁵

Department of Information Technology,
Bengal College of Engineering & Technology, Durgapur

¹Mandavi.shreshtha@gmail.com

²Prasannjit.singh@gmail.com

³Nilotpal_mrinal@yahoo.in

⁴Kalyan.durgapur@gmail.com

⁵S_dg@yahoo.com

ABSTRACT

Heart is one of the vital parts of human body, which maintains life line. In this paper, an efficient composite method has been developed for data compression of ECG signals. ECG waveforms reflect most of the heart parameters closely related to the mechanical pumping of the heart and can therefore, be used to infer cardiac health. After carrying out detailed studies of different data compression algorithms, we used back propagation algorithm to analyse the artificial neural networks. Twelve significant features are extracted from an echocardiogram (ECG). The features of samples are used as input to the neural network. Finally the samples which are used in the database are trained and tested using the Back Propagation Algorithm. The efficiency is observed to be 99.5%. Dual three-layer neural networks with only a few units in the hidden layer are used. It is further observed that input signals are same as supervised signals used in the networks. Back-propagation is used for the learning process.

KEYWORDS

Back propagation, Bipolar coding, Data compression, Echocardiograph Data Set, Neural networks, Linear scaling.

1. INTRODUCTION

The electrocardiogram (ECG) was introduced into clinical practice more than 100 years ago by Einthoven. It provides representation of the electrical activity of the heart over time and is probably the single-most useful indicator of cardiac function. In the cases like critical cardiac patients and ambulatory patients, 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. Also, in an average sized hospital, many tera-bytes of data are generated every year, almost all of which has to be kept and archived. Archiving this large amount of data in the computer memory is very difficult without any compression. Compression methods have gained in importance in recent years in many medical areas like telemedicine, health monitoring, etc. 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. Even though many compression algorithms have been reported so far in the literature, not so many are currently used in monitoring systems and telemedicine.

2. DATA COMPRESSION

Compression is used just about everywhere. All the images we get on the web are compressed, typically in the JPEG or GIF formats, most modems use compression and several file systems automatically compress files when stored, and the rest of us do it by hand. Many compression algorithms exist which 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. Compression rate measures how much the signal can be compressed from the original one. Compression methods used can be lossless and lossy.

2.1. Lossless Compression

Lossless compression implies the original data is not changed permanently during compression. After decompression the original data can be retrieved. The advantage of lossless compression is that the original data stays intact without degradation of quality and can be reused. The disadvantage is that the compression achieved is not very high.

2.2. Lossy Compression

In lossy compression technique, parts of the original data are discarded permanently to reduce file. After decompression the original data cannot be recovered this leads the degradation of quality.



Figure 1. Original image



Figure 2. Compressed Image

3. DATA COMPRESSION TECHNIQUES

Data compression techniques have been classified in a broad spectrum of communication areas such as speech, image and telemetry transmission. The technique of compression used in this paper is explained as follows:-

3.1. Linear Predictive Coding (LPC)

Linear predictive coding (LPC) is defined as a digital method for encoding an analog signal in which a particular value is predicted by a linear function of the past values of the signal. The most important aspect of LPC is the linear predictive filter which allows the value of the next sample to be determined by a linear combination of previous samples. But there is information loss in this technique, thus, it comes under lossy compression.

4. DATA COMPRESSION USING NEURAL NETWORK

A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. The neural networks used in data compression have massively parallel structures and high-degree of interconnections. The compression ratio depends on the ratio of neurons on input layer and on hidden layer. The actual compressed data is obtained from the weights and activation levels of the network. In this paper we have used back propagation technique to train the data set.

4.1. Back Propagation

Back propagation is a systematic method for training multilayer artificial neural networks. It has a mathematical foundation that is strong if not highly practical. It is a multi-layer forward network using delta learning rule, commonly known as back propagation rule.

The training algorithm of back propagation involves four stages:-

- i) Initialization of weights.
- ii) Feed forward
- iii) Back propagation of errors.
- iv) Updating of weights and biases

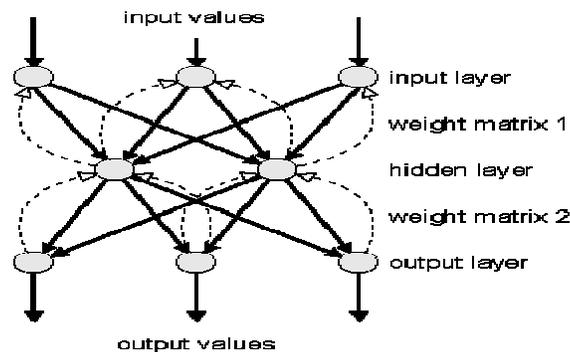


Figure 3. Back propagation neural network

5. DATA SET DESCRIPTION

The database considered in this project has 124 instances and 12 attributes, all of which are numeric-valued. Attribute information are as follows:

Table 1. Description of instances in the data set

S.no	Attribute	Information
i.	Survival	The number of months patient survived (has survived, if patient is still alive). Because all the patients had their heart attacks at different times, it is possible that some patients have survived less than one year but they are still alive. Check the second variable to confirm this.
ii.	Still-alive	a binary variable where '0' indicates dead at end of survival period and '1' means still alive.
iii.	Age-at-heart-attack	age in years when heart attack occurred
iv.	Pericardial – Effusion	a binary variable. Pericardial effusion is a kind of fluid around heart. '0' means no fluid and '1' means fluid.
v.	Fractional-Shortening	a measure of contractility around the heart. Lower numbers indicate abnormal condition.
vi.	EPSS	E-point septal separation, another measure of contractility. Larger number indicate abnormal condition.
vii.	LVDD	left ventricular end-diastolic dimension. This is a measure of the size of the heart at end-diastole. Large hearts tend to be sick hearts.
viii.	Wall-motion-score	a measure of how the segments of the left ventricle are moving.
ix.	Wall-motion-index	equals wall-motion-score divided by number of segments see Usually 12-13 segments are seen in an echocardiogram.
x.	Mult	a derivate variable
xi.	Group	meaningless, ignore it.
xii.	Alive-at-1	Boolean-valued. Derived from the first two attributes where '0' means patient was either dead after 1 year or had been followed for less than 1 year. '1' means patient was alive at 1 year.

5.2. Linear Scaling

The given dataset are in analog form and need to be converted to digital form. Scaling has the advantage of mapping the desired range of variable i.e. ranges between minimum and maximum range of network input. The conversions are based on certain ranges, which are defined for each attribute. There are total twelve attributes. The numerical attributes are in analog form scaled in the range between 0 and 1. The following formulae has been used for linear scaling:-

$$\Delta = X_{\max} - X_{\min}$$

$$Y = \text{Intercept } C = (X - X_{\min}) / \Delta$$

$$\text{Slope (m)} = 1 / \Delta$$

So we can calculate Y for a given X,

$$Y = mX + C$$

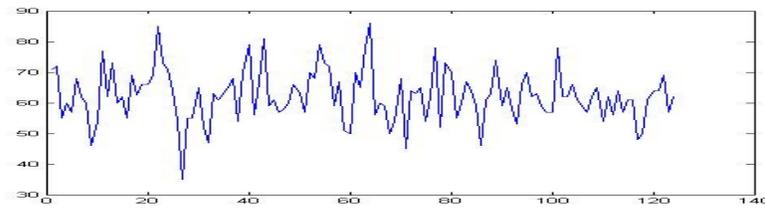


Figure 4. Graph representing one of the attributes of sample analog data

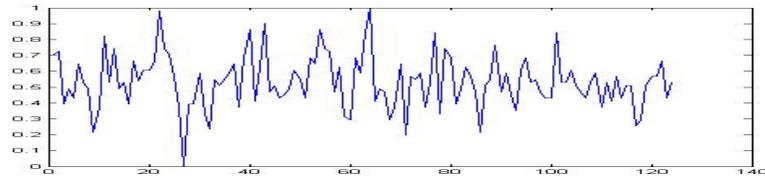


Figure 5. Graph representing linear scaled data

5.2. Bipolar Coding

The numerical attributes are in analog form scaled in the range between 0 and 1. Thus for converting into binary (digital) form, we assign a discrete value of “0” to the attribute value of less than or equal to “0.5”.

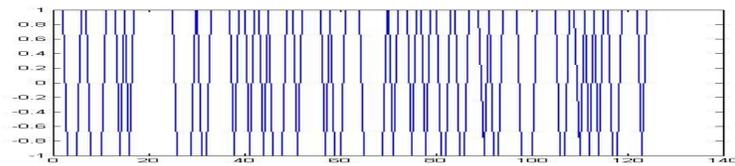


Figure 6. Compressed signal

5.3. Use of Back Propagation In The Data Set Reflected In Graphs

5.3.1. Notations:

- i) Weights: two weight matrices: From input layer (0) to hidden layer (1)
From hidden layer (1) to output layer (2)
Weight from node 1 at layer 0 to node 2 in layer 1
- ii) Training samples: pair of $\{(x_p, d_p) | p=1, \dots, P\}$ So it is supervised learning
- iii) Input pattern: $x_p = (x_{p,1}, \dots, x_{p,n})$
- iv) Output pattern: $d_p = (d_{p,1}, \dots, d_{p,k})$
- v) Desired output: $o_p = (o_{p,1}, \dots, o_{p,k})$
- vi) Error: $l_{p,j} = o_{p,j} - d_{p,j}$ error for output j when x_p is applied.

5.3.2. Pattern classification:

- i) Classification of electric signals
 - Input pattern: 12 features , normalized to real values between 0 and 1
 - Output patters: 3 classes: (First stroke, second stroke, Dead)

ii) Network structure

- 118 input nodes, 3 output nodes
- 1 hidden layer of 3 nodes
- $\alpha = 0.05$ (Learning rate)
- Mean Squared Error: $(0.5S(t_k - y_k)^2)$
- Maximum iteration= 100

6. RESULTS**6.1. Selection Of Learning Rate (α):**

Number of epochs = 100

Table 2. Selection of learning rate

Serial Number	Alpha (α)	Mean Squared Error ($0.5 S(t_k - y_k)^2$)
1	0.9	1.9481
2	0.8	1.7823
3	0.7	0.5663
4	0.6	0.5621
5	0.5	0.54239
6	0.4	0.5300
7	0.3	0.51221
8	0.2	0.4654
9	0.1	0.3211
10	0.05	0.3760
11	0.005	0.3221

Final value of learning rate = 0.05

6.2. Selection Of Momentum Parameter (μ):

Number of epochs = 100

Table 3. Selection of Momentum Rate Factor

Serial Number	Momentum Factor	Mean Squared Error
1	0.9	1.984
2	0.8	0.5821
3	0.7	0.543
4	0.6	0.5321
5	0.5	0.5421
6	0.4	0.5911
7	0.3	0.5992
8	0.2	0.5611
9	0.1	0.3699
10	0.05	0.325
11	0.005	0.3200

Final value of Momentum Parameter (μ) = 0.1

6.3. Test Results for Digital Data

Learning Rate (α) = 0.05, Momentum Parameter (μ) = 0.1, Compression Ratio = 0.974583

Table 4. Test Results for Digital Data

Serial Number	Training Data	Testing Data	Simulation Time (min)	Efficiency (%)
1	20%	80%	3.95	79.55
2	40%	60%	14.78	81.67
3	60%	40%	34.89	83.69
4	80%	20%	47.12	92.17
5	95%	5%	58.07	99.5

6.4. Test Result for Analog Data

Learning Rate (α) = 0.05

Momentum Parameter (μ) = 0.1

Compression Ratio = 0.974583

Table 5. Test Results for Analog Data

Serial Number	Training Data	Testing Data	Simulation Time (min)	Efficiency (%)
1	20%	80%	4.15	72.14
2	40%	60%	14.8	75.43
3	60%	40%	30.79	82.32
4	80%	20%	44.87	85.69
5	95%	5%	49.07	89.57

7. CONCLUSION

Simulation of the back propagation network in this paper has achieved the objective of data compression of ECG signals based on the given data set. Thus, for a supervised input pattern, the output is obtained with a good level of accuracy.

This paper is simulated for the echocardiogram data set. Also, it must be noted that Linear Scaling is used for digitizing the signals and after this process back propagation is applied in order to compress the signals. The tables 2-5 reflect about 99.5% of the accuracy. Hence it can be concluded that back propagation network is best suited for data compression algorithm which proves out to be a lossless compression.

REFERENCES

- [1] <http://archive.ics.uci.edu/ml/datasets.html> for database of Echocardiogram.
- [2] Anuradha Pathak and A. K. Wadhvani, "Data Compression of ECG Signals Using Error Back Propagation (EBP) Algorithm", *International Journal of Engineering and Advance Technology (IJEAT)* ISSN: 2249 – 8958, Volume-1, Issue-4, April 2012.
- [3] R. Rojas, "The Back propagation Algorithm", *Neural Networks*, Springer-Verlag, Berlin, 1996
- [4] Monica Fira and Liviu Goras, "Biomedical Signal Compression based on Basis Pursuit", *International Journal of Advance Science and Research*, Volume-14, January 2012. [5] A. Yilmaz & M. J. English, "Adaptive Non-Linear Filtering of ECG Signals: Dynamic Neural Network Approach, Artificial Intelligence Methods for Biomedical Data Processing".
- [6] Y. Nagasaka, A. Iwata, "Performance Evaluation of BP and PCA Neural Networks for ECG Data Compression. *Neural Networks*", 1993. *IJCNN '93-Nagoya. Proceedings of 1993, International Joint Conference on*, Volume: 1, 25-29 Oct. 1993
- [7] R. Battiti, A. Sartori, G. Tecchiolli, P. Tonella and A. Zorat, "Neural compression: an integrated application to EEG signals, in: *Proceedings of the International Workshop on Applications of Neural Networks*", Stockholm, 1995, pp. 210–219
- [8] *World Congress on Neural Networks*, San Diego: 1994, *International Neural Network Society - 1994 - Psychology - 3580 pages* "Data compression technique using neural networks", June 5-9, 1994
- [9] N. Pradhan, D. Narayana Dutt, "Data Compression by Linear Prediction for Storage and Transmission of EEG signals", *International Journal of Bio-Medical Computing*, Volume-35, Issue-3, April 1994.
- [10] Astola, J., Dougherty, E., Shmulevich, I., Tabus, I. (editors), "Signal Processing", Special issue. on *Genomic Signal Processing*, Vol. 83, No.4, 219 pages, April, 2003\
- [11] C.D. Giurcaneanu, Ioan Tabus, "Escape Sequences for Lossless Audio Compression", *International Symposium on Information Theory and Its Applications*, Sheraton Waikiki Hotel, Honolulu, vol. 1, pp. 386-389, November 5-8, 2000
- [12] Nina F. Thornhill, M.A.A Shoukat Choudhary, Sirish L. Shah, "The impact of compression on data driven process analyses", *Journal of Process Control* 14(2004) 389-398.
- [13] Koch, Karl Rudolf, "Data Compression by multi-scale representation of signals", *Journal of Applied Geodesy*, Volume-5, Issue-1, ISSN-1862-9024, March-2011.
- [14] S.C Saxena, A. Sharma, S.C Choudhary, "Data Compression and Feature Extraction of ECG Signals", *International Journal of Systems Science*, U.K, Volume-28, May-1997.
- [15] J. Chen, S. Itoh, and T. Hashimoto, "ECG data compression by using wavelet transform," *IEICE Trans. Inform. Syst.*, E76-D (12): 1454–1461, 1993.
- [16] Cohen, P. M. Poluta, and R. Scott-Millar, "Compression of ECG signals using vector quantization," in *Proc. IEEE-90 S. A. Symp. Commun. Signal Processing COMSIG-90*, Johannesburg, South Africa, pp. 45–54, 1990
- [17] G. Nave and A. Cohen, "ECG compression using long-term prediction," *IEEE. Trans. Biomed. Eng.*, 40: 877–885, 1993.
- [18] Yaniv Zigel, Arnon Cohen, and Amos Katz, "ECG Signal Compression Using Analysis by Synthesis Coding", *IEEE Transactions on Biomedical Engineering*, 47 (10), 2000.
- [19] J. Cox, F. Nulle, H. Fozzard, and G. Oliver, "AZTEC, a preprocessing program for real-time. ECG rhythm analysis," *IEEE. Trans. Biomedical Eng.*, BME-15: 128–129, 1968.
- [20] S. C. Tai, "Improving the performance of electrocardiogram sub-band coder by extensive Markov system," *Med. Biol. Eng. And Computers*, 33: 471–475, 1995.

AUTHORS

MANDAVI is a B.Tech 3rd year student, Department of Information Technology, Bengal College of Engineering and Technology.



PRASANNJIT is a B.Tech 3rd year student, Department of Information Technology, Bengal College of Engineering and Technology.



NILOTPAL MRINAL is a B.Tech 3rd year student, Department of Information Technology, Bengal College of Engineering and Technology.



KALYAN CHATTERJEE is currently, working as an Assistant Professor, Department of Computer Science Engineering, Bengal College of Engineering and Technology, Durgapur.



Dr. S.DASGUPTA is currently, working as Professor and Head of Department, Computer Science Engineering department, Bengal College of Engineering and Technology, Durgapur. He was also an Ex-scientist at CMERI, Durgapur.

