International Journal For Technological Research In Engineering
Volume 4, Issue 3, November-2016 ISSN (Online): 2347 - 4718

CLUSTERING OF COMPRESSED ECG SIGNAL AS DIGITAL DATA THROUGH SELF ORGANIZING MAP (SOM)

Mayank Mathur
PhD Scholar, Dept. of ECE, Jodhpur National University, Jodhpur

Abstract: In the previous work turning point (TP) algorithm was used to compress ECG signal. Two lead MIT-BIH Arrhythmia Database was compressed and the quality of reconstructed signal is evaluated on the basis of Percentage Root Mean Square Difference(PRD) and Compression Ratio (CR). In this paper the clustering of digital data is done using Self Organized Maps(SOM) . SOM is used to train the input patterns of digital data for different map sizes and window sizes to find suitable parameters for training. The results show that Self Organizing Map is clearly able to cluster the digital data for different types of patterns.

I. INTRODUCTION

The electrocardiogram (ECG) is the graphical recording of the time varying voltages produced by the myocardium during the cardiac cycle. The instrument used to obtain and record the electrocardiogram is the electrocardiograph. Figure 1 shows the basic waveform of the normal electrocardiogram.

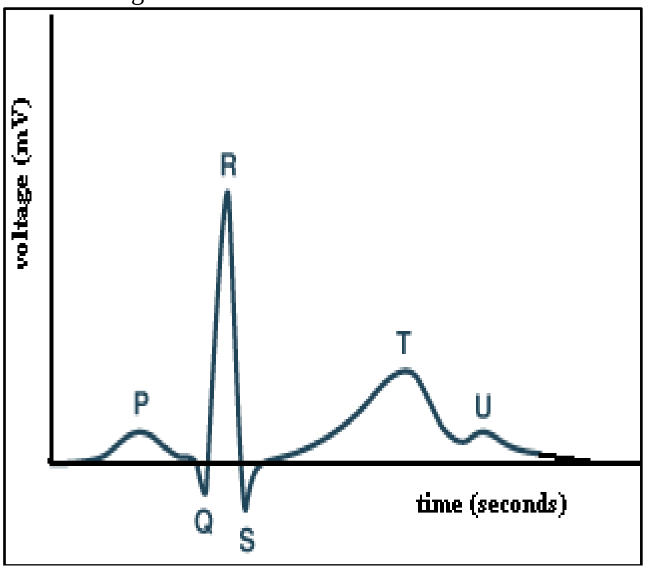

Figure 1 the Basic Electrocardiogram (ECG)

From this topic ECG would be used for electrocardiogram. The main components of the ECG waveform the P, QRS and T waves reflect the rhythmic electrical depolarization and repolarization of the myocardium associated with the contractions of the atria and the ventricles. The electrocardiogram is used clinically in diagnosing various diseases and conditions associated with the heart. The ECG consists of various peaks/complexes, intervals and segments which are shown in Figure 1.4.

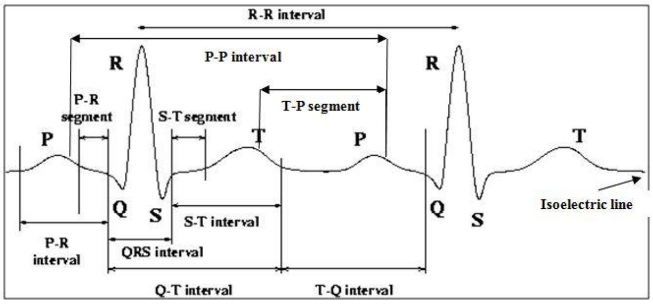

Figure 2The Electrocardiogram with its peaks/complexes, intervals and segments

(A) Isoelectric Line- It is the baseline which indicates that electrical charges on either side are balanced.

(B) ECG Peaks/Complex

- P peak- It is a rounded deflection in ECG trace which represents the atrial depolarization.
- Q peak- It is the initial downward (negative) deflection, related to the initial phase of depolarization of the ventricular myocardium, the depolarization of the interventricular septum.
- R peak- It is the initial upward deflection of the QRS complex, following the Q wave in the normal electrocardiogram and representing early depolarization of the ventricles.
- S peak- It is the downward deflection of the QRS complex following the R wave in the normal electrocardiogram and representing late depolarization of the ventricles.
- T peak- It is the deflection produced by ventricular repolarization.

QRS Complex- The QRS complex is the electrical wave that signals the depolarization of the myocardial cells of the ventricles.

(C) ECG Segments

- PR Segment- It is the isoelectric tracing that follows the P wave and ends with the deflection of the Q wave. It represents the delay of the electrical impulse at the atrioventricular (AV) node.
- ST Segment- It is essentially a period of diastole for the heart and represents the period from the end of systole to the beginning of repolarization of the ventricles.
- TP Segment- It is the isoelectric interval on the electrocardiogram (ECG). It is the region between the end of the T wave (ventricular repolarization or electrical inactivation) and the next P wave (atrial depolarization or electrical activation).

(D) ECG Intervals

- RR Interval- It is the interval from the peak of one QRS complex to the peak of the next as shown on an electrocardiogram. It is used to assess the ventricular rate.
- PP Interval- It is the time from the beginning of one P wave to that of the next P wave, representing the length of the cardiac cycle.
- PR Interval- It is the time delay between the atrial repolarization and ventricular activation.
- QT Interval- It is the total time for ventricular depolarization and repolarization.
- QRS Interval- It is the length of time between ventricular depolarization (the Q wave) and repolarization (the T wave).

II. KOHONEN'S SELF ORGANIZING MAP

Self Organizing Map .i.e. SOM is a sheet-like artificial neural network, the neurons of which become specifically tuned to various input signal patterns or classes of patterns through an unsupervised learning process. The input patterns are of high dimensionality but the neurons in the maps are placed on the lattice of 1 or 2 dimensions. All the neurons in the map differently respond to each input pattern applied to them, thereby they become feature selective. All the neurons in the map differently respond to each input pattern applied to them. As per competition learning rule, only one output neuron wins the competition, that neuron is called the winning neuron or winner-takes-all neuron. All neurons have their associated weights and position in the map space. So the locations and weights of all neurons that are tuned in the learning phase are ordered with respect to winning neuron in such a way that meaningful coordinate system for every input pattern is created over the lattice. Figure 3 shows the input vectors, the connections and synaptic weights between the input vectors and the output neurons of the map.

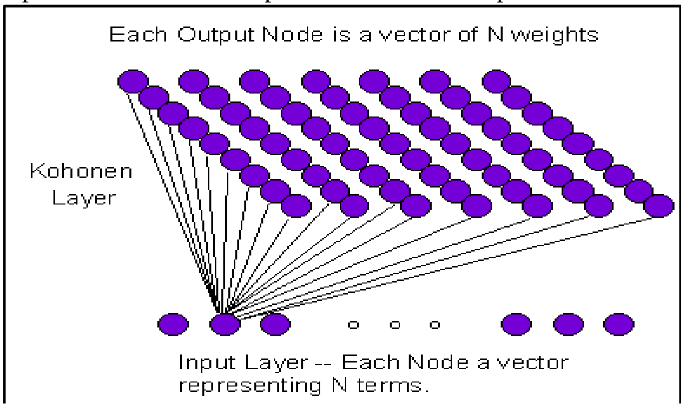

Figure 3 Inputs and Weights in SOM

There are four basic steps of the SOM algorithm; initialization, sampling, matching and updating, in which later three are repeated for number of iterations up to the formation of the feature map.

Initialization: All the synaptic weights $\mathrm{m_j}(0)$ for $\mathrm{j} = 1,2,\ldots,\mathrm{l}$ are initialized randomly by using a random function generator where l is the number of neurons in the network.

Sampling: A sample input data is chosen randomly from the defined region. This is activation pattern in the map, therefore is applied to all the neurons of the network.

Matching: All neurons are processed for calculation of the distance function through which the Best Matching Unit c(x) is determined. Following formula is used for this determination,

$$\mathrm{c(x)} = \arg\min_{\mathrm{j}} \mathrm{d_j(x)} \quad \text{for } \mathrm{i} = 1,2,\ldots,\mathrm{l}$$

Cooperation: After the winning neuron has been detected, the neighborhood of the winning neuron is found out. This neighborhood consists of the neurons which are nearer to the winner. So the cooperation provided by the winning neuron to the other neurons in the network is more to nearer neurons than the neurons farther away.

Updating: The synaptic weight of all the neurons are updated with respect to the winning neuron and in accordance with the neighborhood function $\mathrm{e_{j,i(x)}(n)}$ centered at the winning neuron c(x) as,

$$\mathrm{m_j(n+1)} = \mathrm{m_j(n)} + \alpha \mathrm{e_{j,i(x)}(n)}(\mathrm{x} - \mathrm{m_j})$$

where, α is the learning rate parameter whose initial value is set to 0.1.

Continuation: The process is repeated from step 2 until no identifiable changes in the feature map are observed.

III. PARAMETERS USED IN THE SOM ALGORITHM

The parameters which are needed to be taken before the process of learning of the SOM begins are:

1. Input Data
2. Dimension of the map
3. Initial Weight Vector
4. Stopping Criterion

1. Input Data

The input data that is presenting to Self Organizing Map (SOM) is a digital (binary) data. Every input pattern is taken to be consisting of 8 elements, thereby forming 28= 256 patterns. Thus, the data would be presented to network in the following form as:

$$x[n] = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8]^T$$

where, n varies from 1 to 256.

2. Dimension of the Map

In present research work This thesis works on the 2 dimensional map whose two dimensions are taken in row and col where row is the size and col is the column size respectively where whereas the no. of elements of the weight vector of neurons or in the input pattern is taken in p. The values of $row \times col$ are all assumed to be odd.

3 Initial Weight Vector

Every neuron in the map will be receiving input from total of 256 input patterns and as each pattern consists of 8 elements, the neurons will be excited by 8 elements, thereby contributing 8 synaptic weights for each neuron.

But for the inception of the process of Self Organizing Map, the weights of all the neurons are initialized to a value that lies between 0 and 1 using a random weight vector. Thus for every neuron there will be 8 weights .i.e., 8 input patterns linked to it. So if there is a 7×7 map then total neurons in

International Journal For Technological Research In Engineering
Volume 4, Issue 3, November-2016 *ISSN (Online): 2347 - 4718*

the network are 49 and total weights would be 49*8 = 412.

4. Stopping Criterion

Theoretically, the stopping criterion for the self organizing map algorithm is to check whether the weights have saturated or not. If the weights have saturated then the algorithm has to be stopped. But practically this was not adopted so the loop has been iterated a number of times that is predefined. Initially, for the map to learn 1 Lac (1, 00,000) iterations are taken and that is reduced to 80,000 based on other parameters. So, the map learns for the input 1 Lac times.

IV. LEARNING PHASE OF SOM

Learning phase of the SOM starts with the presentation of the input pattern that is randomly selected from the 256 input patterns stored in a file. Following steps are performed with the input data:

1. The Best Matching Unit (BMU) is found for the input pattern presented to the network. This is also called the winner or the winning neuron for that input pattern. This winner is selected from the bunch of neurons by using the Euclidean distance phenomena so that the winner selected will be having the minimum Euclidean distance from the input data corresponding to the maximum similarity of winner.
2. After selecting the winner, this winner makes up its neighborhood comprising of the exciting neuron. These neurons are selected again on the basis of Euclidean distance but with respect to the winner. All the neurons are assigned a value termed as neighborhood function which is maximum for the winner and decreases with the increasing distance of the neurons from the neighbor.
3. Depending on the values of the neighborhood function, eta and the sigma values, the weights of the neurons are updated using equation (6.17).

$$m_j(n+1) = m_j(n) + \alpha(n) e_{j,i(x)}(n)\big(x(n) - m_j(n)\big)$$

This finishes the learning phase of the Self Organizing Map Algorithm. Thus a map has been trained and is now ready for the analysis of the data.

V. TESTING PHASE

1. Image of the weights of neurons

The weights are randomized between 0 and 1. After learning the network of neurons, an image using the 'imagesc' function implemented on sum of all weight elements in a weight vector of a neuron.

The initial parameters are taken as follows:

$row = 5 ; col = 5 ; p = 8 ;$

$row = 7 ; col = 7 ; p = 8 ;$

$row = 9 ; col = 9 ; p = 8 ;$

$row = 11 ; col = 11 ; p = 8 ;$

This image is shown in Figure 4 which shows a 2 dimensional image map of weights of each neuron. Here every weight element is summed up to form a single weight of each neuron. Following figure shows variation of this color map for different sizes of the map.

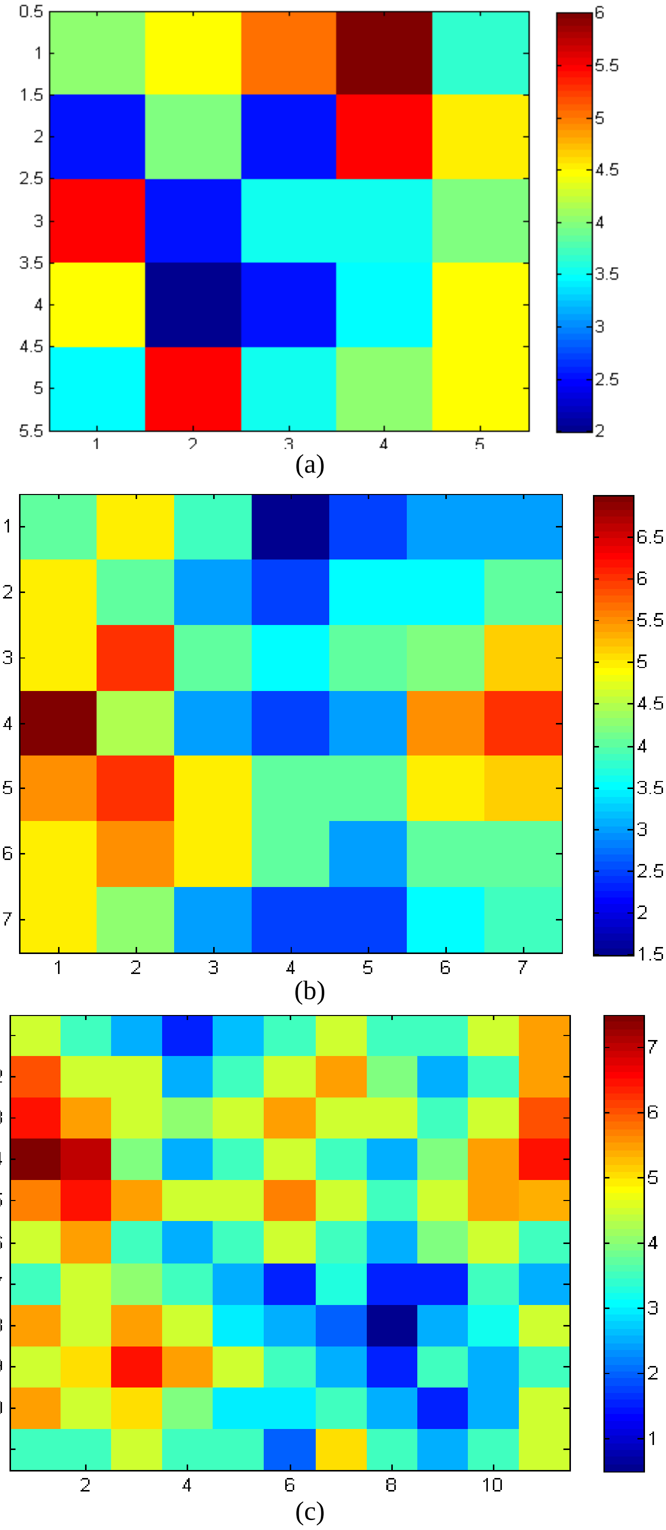

Figure 4 (a) A 5x5 map (b) A 7x7 map (c) A 11x11 map shows the color image of the weights of each neuron when their all elements of weight vector are summed up.

The figure 4 shows clearly the distinction between the patterns mapping on neurons of different dimensions. These figures are interpreted in terms of the similar information mapped on the nearby neurons. Following are the conclusions that can be drawn based on the Figure 7.1.

VI. LEARNING RATE PARAMETER

The learning rate parameter σ(n) plays a vital role in determining the behavior of the algorithm. This parameter is given by Equation (7.8),

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_2}\right)$$

which shows that it depends on the initial value of learning rate parameter, time and the constant value of tau τ_2.

The choice of learning rate parameter is independent. As the learning at the beginning of the process should be high so the maximum value of 1 is assigned to it .i.e., σ_0=1. The values of tau has been experimented and a reasonable value of tau has been choosen. The variation of learning rate parameter with varying values of tau is shown in Figure 5.

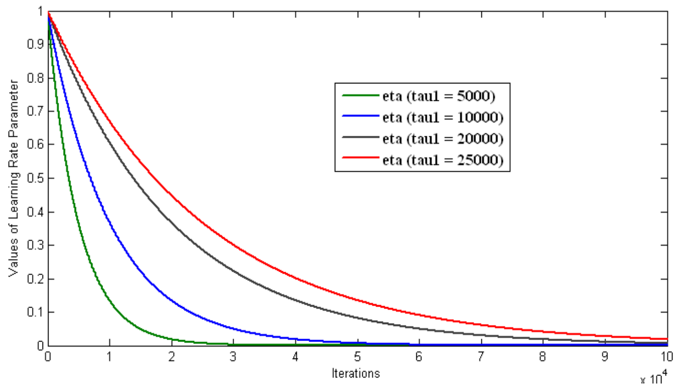

Figure 5 Variation of Learning Rate Parameter with variation of τ_2

The values of tau has been taken as the multiple of 500 to obtain an influential changes in the eta. Taking first τ_2=5000, it shows that the learning decreases exponentially at a faster rate. Due to this, the weights do not converge properly which escapes the ability of the algorithm to make them equal to the average of input patterns. When the algorithm is treated with τ_2=10000, the same problem was coming. With τ_2=20000 and τ_2=25000, we were getting the same results, the weights are reaching their proper values. But with only difference that with τ_2=25000 more time is taken to reach the same results. So the optimum value of tau achieved is τ_2=20000. This also fixed the total number of iterations of the process. We then take 100000 iterations for the process to learn the input patterns.

VII. WIDTH OF THE NEIGHBORHOOD (SIGMA)

The width of the neighborhood function is large at the start of the algorithm and then it decreases exponentially with time. It is given by Equation (7.8)

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_2}\right)$$

whereα_0 is the initial value of width of neighborhood and which is taken as (row-1)/2. It decreases with time depending on the value of τ_2 which is given by-

$$\tau_2 = \frac{30000}{(\log \sigma_0 + 1)}$$

When the dimension of the map is 7x7, the value of α_0=3, then τ_2 is 1.4295e+004. Variation of sigma with time can be seen in the Figure 6.

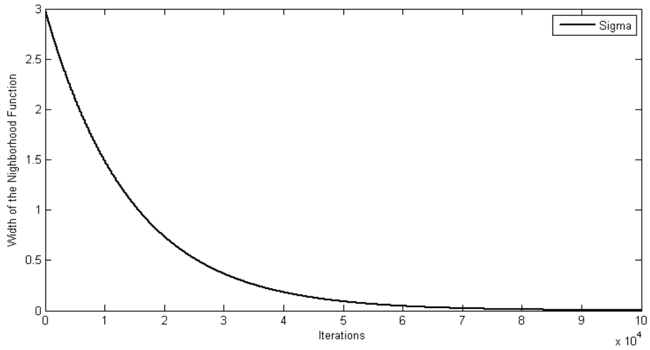

Figure 6 Variation of Sigma

VIII. CONCLUSION

In addition to that clustering of digital data is done using Self Organized Maps(SOM) . SOM is trained input patterns of digital data for different map sizes and window sizes to find suitable parameters for training. The results show that Self Organizing Map is clearly able to cluster the digital data for different types of patterns. Similar kind of digital data is clustered into nearby region of SOM. This shows the clustering capability of self organizing maps which preserves the similar characteristics of a pattern and similar patterns are mapped on nearby regions.

REFERENCE

[1] Boonyaves, P., Paisalsing, P., Totarong, P., Jitapunkul, S. (2004b) "Performance evaluation of multiquadric interpolation technique for ECG signal compression" IEEE Conference Region TENCON 2004, Proceedings, pp 431- 434

[2] Cetin, A. E., Köymen, H. “Compression of Digital Biomedical Signals.” The Biomedical Engineering Handbook: Second Edition.

[3] Cherkassky, V., Kilts, S. (2001) "Myopotentialdenoising of ECG signals using wavelet thresholding methods" Neural Networks, vol. 14, no. 8, pp 1129-1137

[4] Gilmour, R.F. (2004) "The anatomy of an arrhythmia" The Journal of Clinical Investigation, vol. 113, no. 5, pp- 662-664

[5] J. I. Mwasiagi, “SELF ORGANIZING MAPS APPLICATIONS AND NOVEL “,Edited by JosphatIgadwaMwasiagi

[6] L. Owsley, L. Atlas, and G. Bernard, “Self-organizing feature maps with perfect organization,” 1996 IEEE Int. Conf. Acoust.Speech, Signal Process. Conf. Proc., vol. 6, pp. 3557–3560.

[7] Li, H. et al. Genetic algorithm for the optimization of features and neural networks in ECG signals classification. Sci. Rep. 7, 41011; doi: 10.1038/srep41011 (2017).

[8] Sandham, W.A., Thomson, D.C., Hamilton, D.J. (1995), "ECG compression using artificial neural networks" IEEE Conference on Engineering in Medicine and Biology, Proceedings, vol. 1, pp 193-194

[9] T. Villmann, R. Der, M. Herrmann, and T. M.

Martinetz, “Topology preservation in self-organizing feature maps: exact definition and measurement.,” IEEE Trans. Neural Netw., vol. 8, no. 2, pp. 256–66, Jan. 1997.

[10] V. R. Kumar, A. Sivanantharaja, "Feed Forward Neural Network Optimized Using PSO and GSA for the Automatic Classification of Heartbeat", Middle-East Journal of Scientific Research, vol. 23, no. 5, pp. 896-901, 2015.

[11] Waqas Ahmed, Shehzad Khalid,” ECG signal processing for recognition of cardiovascular diseases: A survey “, Sixth International Conference on Innovative Computing Technology (INTECH), 2016, IEEE Xplore: 09 February 2017 DOI: 10.1109/INTECH.2016.7845089

