Indian Journal of Science and Technology, *Vol 10(13), DOI: 10.17485/ijst/2017/v10i13/104487, April 2017*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Heart Rate Variability using Neural Network

Munish Suri and Eshaan Verma

Department of Electronics and Communication, AIACTR, Delhi – 110031, India;
munishsuri95@gmail.com, eshaan@cetpainfotech.com

Abstract

Objective: This paper investigates heart rate variability using neural networks. **Methods:** A software is developed to detect the heart rate variation using ECG signals. For this purpose the signal is divided into sub samples that are overlapping to a certain extent. The signals are transformed into the frequency domain using FFT (Fast Frequency Transform) and to decompose the signal, wavelet transform is used. Further the features are extracted using the above transforms which are fed into the neural networks**. Findings:** Heart rate variation is calculated using neural networks where the features are used to train them. For training, two learning algorithms are used, LM (Levenberg Marquardt) and BR (Bayesian Regularization). LM is found to converge faster than BR but the latter has higher efficiency. **Application:** Variation in heart rate can be used for better detection of diseases.

Keywords: Bayesian Regularization, Heart Rate Variability, Levenberg Marquardt Machine Learning, Signal Processing

1. Introduction

The purpose of this research work starts with a fresh zeal of interest in the complete details of the ECG signals along with the introspection of the level of accuracy so obtained and then expanding the idea from paper to software. Selection of software is a very critical issue in this regard and is a very challenging task that leads to a good level of idea incorporated from the feature extraction technique and then recognition. The ECG signals of patients are used to recognize the Heart Rate Variability (HRV). HRV is the physiological phenomenon of variation in the time interval between heartbeats. It is measured by the variation in the beat-to-beat interval also known as R-R intervals. Electrocardiogram (ECG) is the electric signal originating from heart. The complete work finally comes down to identifying the percentage recognition accuracy.

The prime focus of this research is to identify the heart rate variability in patients. To conduct this research a fresh setup is created which uses software facilities rather than any kind of hardware facility so that a subject with any sort of problem can be identified easily. But the biggest challenge is to use a good dataset that helps in creating an initial model from which classification, prediction can be done more effectively. Thus the race of this challenge begins with using a database available from physionet website and then taking it to a level of recognition by using different types of classifiers like support vectors, discriminants and others.

HRV is measured as the time gap between the heart beats that vary as we breathe in and out. For the heart to beat, the SA node (Sinoatrial node) known as the natural pacemaker sends out an electrical pulse due to which the upper heart chamber contracts. The AV node (Atrioventricular node) then sends an impulse into the ventricles and the lower heart chamber contracts. The SA node sends another signal to the atria to contract, which starts the process again.[1]

Its analysis requires signal acquisition, detection of beat to beat and quality control of an ECG signal. ECG signals can be classified according to the length i.e. long term recordings and short term recordings. In short term the sample usually ranges between 2-5 minutes whereas in long term it usually lasts 24 hours.

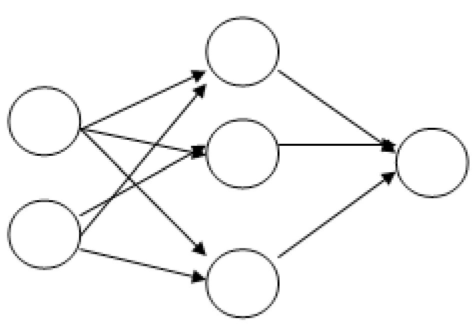
In this paper we have used neural networks[2] which use different algorithms for attaining the co-efficients and for the recognition of heart rate signals, using pre defined classes of the signals. For this we transform our signal

**Author for correspondence*

using FFT (Fast Fourier Transform) to attain the values in frequency domain and we use Wavelet transform to attain properties such as entropy and energy. Usually HRV is calculated using various parameters such as SDNN (Standard Deviation of the NN interval), SDANN (Standard Deviation of the NN Average interval), RMSSD (Root Mean Square of the Successive Differences of NN intervals) and NN50 i.e the number of interval differences of successive NN intervals greater than 50 ms, in frequency domain analysis. In frequency domain analysis we use PSD (Power Spectral Density) i.e how power is distributed as a function of frequency. In this experiment there are basically two phases, first is feature extraction and second is the identification using neural network.

2. Neural Networks

Neural networks are like brain analysis, in which we train our computer using a database so that when it encounters anything (signals in this case) related to our data then it easily identifies it. A neural network is shown in Figure 1 where there are input units, one hidden layer and an output layer. The input layer is connected to the hidden layer which is further connected to the output layer. They are connected using weights and we have to optimize those weights so as to attain the desired output. We use Rule, which optimizes our weights. It has wide applications such as anomaly detection, signal processing, pattern recognition, image processing etc.[3]



Input Layer Hidden Layer Output Layer

Figure 1. Neural Network.

Use of neural networks has many advantages such as lesser data requirement during processing. ANN with Back propagation (BP) learning algorithm is widely used in solving various classifications and forecasting problems. Even though BP converges and is slow but it is guaranteed.[4]

3. Algorithms For Neural Networks

3.1 Levenberg–Marquardt

It is a backpropogation algorithm and uses the advantages of both Gradient descent and Gauss Newton method. In Gradient descent, as shown in Eq (1)

$$x_{i+1} = x_i - \lambda \nabla f \qquad (1)$$

We use scaled gradient to find out the minima. But this step takes a large number of iterations and it converges very slowly. The error in this case is usually high while minimizing the cost function, and also the curvature of error is not the same everywhere.[5-7]

This algorithm is suitable for small and medium sized data, as it is stable and converges fast.

In case of Gauss – Newton method the gradient is calculated as shown in Eq (2)

$$\nabla f(x) = \nabla r(x) r(x) = J(x)^{T} r(x) \qquad (2)$$

where $J(x) = \nabla r(x)$ T is the Jacobian of $r(x)$, i.e.

$$J(x) = \left\{ \begin{array}{ccc} \frac{\partial r_1(x)}{\partial x_1} & \cdots & \frac{\partial r_1(x)}{\partial x_n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ & \cdot & \\ \frac{\partial r_m(x)}{\partial x_1} & \cdots & \frac{\partial r_m(x)}{\partial x_n} \end{array} \right\}$$

The hessian is calculated as:

$$\nabla^2 f(x) = J(x)^T J(x) + Q(x) \qquad (3)$$

For Gauss Newton method Q(x) is 0. After combining the above equation we get Eq (4)

$$x_{i+1} = x_i - (J_k^T J_k)^{-1} J_k^T e_k \qquad (4)$$

where e_k is the error matrix,. In Levenberg Marquardt algorithm, the invertibility of the hessian is sure .

$$H = J_k^T J_{k} + \mu I \qquad (5)$$

where μ is a positive number and I is the identity matrix. Hence the expression becomes

$$x_{i+1} = x_i - (J_k^T J_{k} + \mu I)^{-1} J_k^T e_k \qquad (6)$$

3.2 Bayesian Regularization

They are robust networks, removing the process of cross – validation. For this network we need to define the network and the prior distribution of the weights. The parameter □ is unknown and let it be defined according a distribution say normal. For this we need to calculate the conditional probability as shown in Eq (7) where y is the output and x is the input. In[8-10] Bayesian regularization minimizes the linear combination of squared errors and weights. It also modifies the linear combination so as to give good generalization qualities in the end.

$$P(\square/S) \;\square\; \Big(\prod_{i=1}^{m} P(y^i/x^i, \square)\Big)\, P(\square) \qquad (7)$$

Hence, to make a new prediction on x we require the conditional probability

$$P(y/x,S) \;=\; \int P(y/x, \square)\, P(\square/S)\, d\square \qquad (8)$$

$$E[y/x,S] \;=\; \int y\, P(y/x, S)\, dy \qquad (9)$$

We use maximum likelihood function for the error measurement.

$$\mathrm{Max} \sum_{i=1}^{m} \mathrm{Log}(P(y^i/x^i, \Theta)) \qquad (10)$$

3.3 One Step Secant

BFGS requires more storage and complexity. Thus there is a need for secant approximation for computation. This algorithm does not store the Hessian matrix and the new search direction can be calculated without calculating the matrix inverse. It also reduces the complexity from O (N X N) to O (N) .There is a slow convergence in this method and some second order information is also lost during the process.[11]

4. Feature Extraction

This is the most important phase of this experiment. It is not possible to train our network using such wide values as there are issues of memory and computation time. We take the signal and transform into various domains, unlike the standard parameters, we use wavelet transform and then compute the parameters.

In time domain analysis we use mean, median and variance, where mean is the measurement of the central tendency, median corresponds to the middle value and variance is the expectation of the squared deviation of the random variables from the mean.

For measurement in the frequency domain we use FFT (Fast Fourier Transform). We could have used DFT (Discrete Fourier Transform) in which

$$\mathbf{X[k]} = \sum_{n=1}^{N} x[n]\, e^{-j2\pi kn/N} \qquad (11)$$

where x[n] is the discrete value in time domain, N is the total size and X[k] is the output in the frequency domain. However when the size of data is big then this algorithm becomes highly complex and takes more time hence we require FFT to minimize the number of steps and the computational time.[12]

Discrete wavelet transform is also used for the decomposition of the signal. There are many short waves (wavelets) available but we have used 'haar' wavelet in this case. The effect of this shifting and scaling process is to produce a time-scale representation. To represent dilation we can use a tree of low pass and high pass filters and the original signal is decomposed into lower resolution components. The maximum number of dilations that can be performed is dependent on the input size of the data to be analysed, with 2N data samples enabling the breakdown of the signal into N discrete levels using the discrete wavelet transform.[13–14]

Figure 2 shows the input signal is 8B.The 8B signal is passed through a series of High Pass (HP) filters and Low Pass (LP) filters to decompose into 2 B signals.

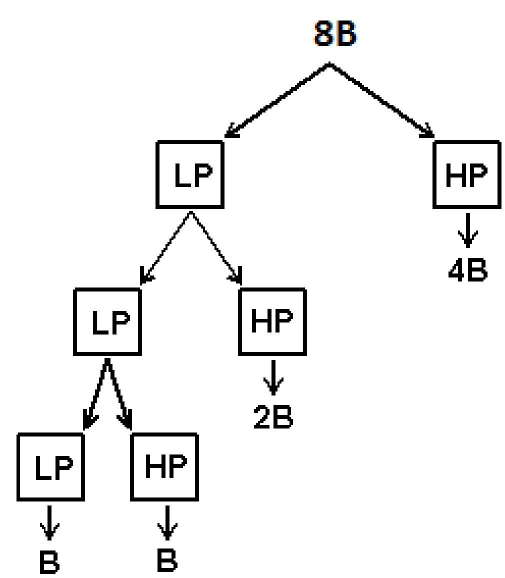


Figure 2. Decomposition of 8B Signal Using Wavelet Transform.

In case of ‘haar’ wavelet there is a mother function as well as a scaling function.

The mother function:

$$\Psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} < x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

And

Scaling function:

$$\Psi(x) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

The properties that are being used after wavelet transform:

4.1 Entropy

It is the measurement of the degree of randomness of the signal. Higher the entropy, greater are the chances of error.

$$En_i = \sum_{k=1}^{N} \mathbf{C}_{ij} \log(\mathbf{C}_{ij}^2) \tag{14}$$

4.2 Energy

Energy of the signal is defined as:

$$En_i = \sum_{k=1}^{N} |\mathbf{C}_{ij}|^2 \tag{15}$$

where C_{ij} is the discrete value of i^{th} row and j^{th} column .

5. Experiment

Figure 3 shows the experiment is done using sample signals in the given dataset with ECG signals. In this approach the signals have been divided into sub samples on basis of their properties. The typical division is into 8 sub samples with overlapping. These samples are fed into our network for training and testing purpose, 50% of them are used for training and rest for the testing. 3 different networks which are trained using different algorithms are used. When testing signals are fed into the network we check the overlapping of the signals with others. Basically features that were extracted in the previous phase are used.

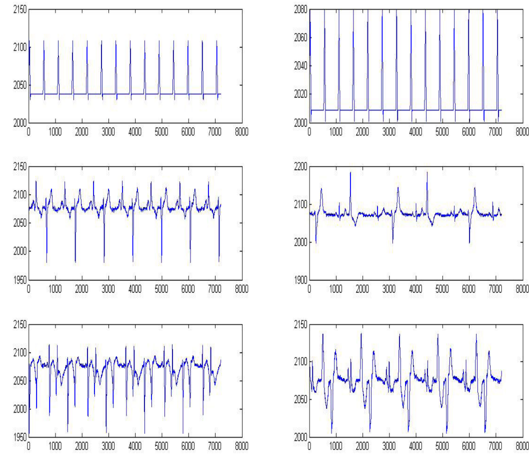


Figure 3. Sample ECG Signals.

The percentage efficiency is calculated using this information, which is shown in Table 1. If the overlapping percentage is high then there are less chances of heart rate variation i.e. HRV (heart rate variability) is inversely proportional to the percentage of overlapping.

Table 1. Identification of sub samples

Signal no	BR (Bayesian Regularization)	LM (Levenberg Marquardt)	OSS (one step secant)
1	4	4	4
2	4	4	4
3	3	2	0
4	3	3	2
5	4	3	2
6	2	1	2
7	4	4	4
8	4	4	4
9	4	4	0
10	4	4	4
Total	36	33	26

The signals after feature extraction are shown in Figure 4, 5, 6.

Figure 4 represents the signals with overlapping in time domain and properties such as mean and variance. Figure 5 represents energy calculated after wavelet transform and Figure 6 represents the energy after FFT (Fast Fourier transform).

After this stage, the network is trained and overlappings are calculated. As shown in Table 1, each signal is divided into 8 samples -4 for training and 4 for testing.

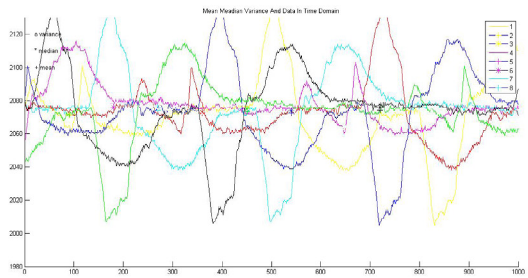


Figure 4. Time Domain Representation of Signals and their Properties.

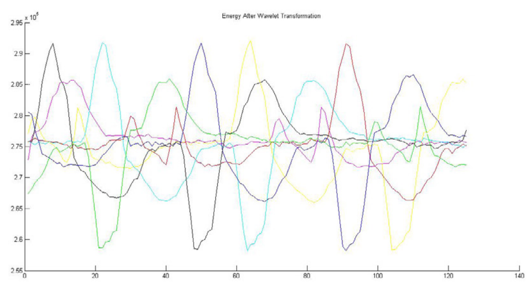


Figure 5. Energy After Wavelet Transform.

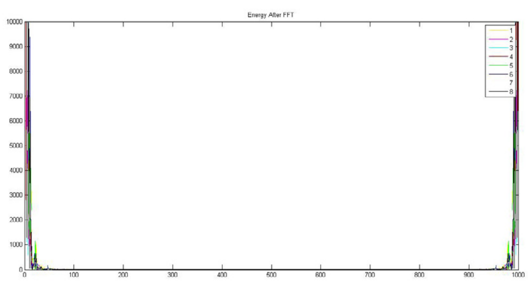


Figure 6. Energy after FFT.

6. Conclusion

From the data given in the tables, it is clear that BR performs very well on this set and finds out the variation in 6 ,3 ,2 signal number whereas LM is less efficient and OSS is the least. The thing to be noted in this case is that the performance time taken in the case of LM is less as it converges very fast. In the above case, the samples detected are 36/40, 33/40 and 26/40 for BR, LM and OSS respectively.

Therefore, if our machine is not bound to time constraints then we may use BR (Bayesian Regularization) else LM (Levenberg Marquardt) is the preferred option.

7. References

1. Christoph B, Kurt S, Stijn WD, Leonhardt S. Adaptive Beat-to-Beat Heart Rate Estimation in Ballistocardiograms. IEEE Transactions on Information Technology in Biomedicine. 2011; 15(5): 778–86. Crossref PMid:21421447
2. A Brief Introduction to Neural Networks. David Kriesel dkriesel.com.27/05/2005.
3. Neural Networks. http://natureofcode.com/book/chapter-10-neural-networks/. Date Accessed: 14/12/2009.
4. Davis Artificial Intelligence Meetup http://www.ijritcc.org/download/Research%20Paper%20on%2 Basic%20of%20Artificial%20Neural%20Network.pdf. Date Accessed: 15/11/2016.
5. Hasan Muaidi. Levenberg-Marquardt Learning Neural Network For Part-of-Speech Tagging of Arabic Sentences. Wseas Transactions On Computers. 2014; 13: 300–09.
6. Levenberg–MarquardtTraining.http://www.eng.auburn.edu/~wilambm/pap/2011/K10149_C012.pdf. Date Accessed: 25/02/2010.
7. Saini LM, Soni MK. Artificial Neural Network based Peak Load Forecasting using Levenberg-Marquardt and quasi-Newton methods. IEEE Proceedings General Transmission Distribution. 2002. Crossref
8. Ji Q, Zhang L. A Bayesian Network Model for Automatic and Interactive Image Segmentation. IEEE - IEEE Transactions on Image Processing. 2011; 20(9):2582–93. Crossref PMid:21356618
9. Vaart V. Bayesian Regularization. International Congress of Mathematicians. Hyderabad: 2010 Aug.
10. Zhu J, Chen N, Xing EP. Bayesian Inference with Posterior Regularization and Applications to Infinite Latent SVMs. Journal of Machine Learning Research. 2014; 15: 1799–1847.
11. Battiti R. First And second order for learning: between steepest descent and newton's method. https://www. researchgate.net/prof ile/Roberto_Batt it i / publication/2498372_First_and_SecondOrder_Methods_for_Learning_Between_Steepest_Descent_and_Newton's_Method/links/5537b11c0cf2058efdeae0b6.pdf.
12. Signal Processing. http://www.wavemetrics.net/doc/igorman/III09%20Signal%20Processing.pdf.
13. Dhiman R, Saini JS, Priyanka . Genetic Algorithm Tuned Expert Model for Detection of Epileptic Seizures from EEG Signals. Applied Soft Computing. 2014 June; 19: 8–17. Crossref
14. Mallat SG, Zhong S. Characterization of Signals from Multiscale Edges. IEEE Transactions Pattern Analysis and Machine Intelligence. 1992 July; 14(7): 1–23. Crossref