Impact of Missing RR-interval on Non-Linear HRV Parameters

Manjit Singh
Dept. of Electronics and
Communication
Guru Nanak Dev University
Regional Campus
Jalandhar, India

Butta Singh
Dept. of Electronics and
Communication
Guru Nanak Dev University
Regional Campus
Jalandhar, India

Ashwani Rajput
Dept. of Electronics and
Communication
Guru Nanak Dev University
Regional Campus
Jalandhar, India

ABSTRACT

The impact of missing RR-interval data on nonlinear heart rate variability (HRV) analysis with and without interpolation were investigated. In this study, randomly selected data (with frequency of 5 samples up to 50) were removed from actual data (taking first 1000 samples) in the MIT-BIH arrhythmia RR interval database of 10 subjects having 1000 sample data points in each set. In all, the tachograms the artefacts are removed first from the 1000 samples. Poincare plot and entropy analysis were executed for the nonlinear HRV parameters, and the absolute relative errors between the data with and without the missing data duration for these parameters including the interpolation were calculated. In this process, the usefulness of reconstruction was considered when there is missed rr-interval, for which several interpolation methods (linear, delete, and zero order interpolation) were used and the best interpolation method having less error in the HRV analysis was chosen. During the work and performing all the interpolation methods, the delete interpolation gives best results for the reconstruction of data while analysing the HRV non-linear parameters.

Keywords

HRV Analysis, Poincare Plot, Entropy, Missing data, Interpolation

1. INTRODUCTION

Heart rate variability (HRV) parameters can be affected by missing RR-interval data which can be produced by motion artifacts or other electronic or mechanical external stimuli. Previous works in this field has compared the frequency domain, time domain and non-linear domain of the HRV with and without reconstruction [1]. In this work, representative nonlinear approaches to HRV analysis (Poincare plot and entropy analysis) are investigated in order to understand the effect of missing data and reconstructed data through interpolation. The non-linear parameters were studied as they are very essential tools which depicts the irregularities graphically and also due to the ease of finding those irregularities on graph. In these simulations, several reconstruction methods for missing RR-interval data are used and the best method for each nonlinear parameter is selected [2].

This paper is organized as follows: definitions of the nonlinear HRV parameters used in this study and the clinical research for each parameter are briefly described in Section 2. In Section 3, explanations are offered for the processes by which the RR-interval data used for the missing data interpolation was obtained. In section 4, the methodology and database used is discussed and flowchart is provided for the

methodology used and then the results for each nonlinear parameter are compared. Finally, the impact of missing RR-interval data is thoroughly discussed in Section 5.

2. NONLINEAR HRV PARAMETERS

2.1 Poincare plot analysis

Poincare plot analysis (PPA) method is a useful visual tool to assess the dynamics of HRV by representing the characteristics of time series fluctuation, a technique which is based on nonlinear dynamics. In this form of geometrical analysis, the successive values of RR-interval data are scattered as points for each pair [3]. The standard deviation along the major axis (SD2) of the data and the standard deviation along the minor axis (SD1) can be obtained by the following calculations:

$$SD1 = \sqrt{Var(\sqrt{\tfrac{1}{2}}RRn - \sqrt{\tfrac{1}{2}}RRn + 1)} \qquad \ldots 2.1$$

$$SD2 = \sqrt{Var(\sqrt{\tfrac{1}{2}}RRn + \sqrt{\tfrac{1}{2}}RRn + 1)} \qquad \ldots 2.2$$

where RRn is a time series of RR-interval data with a heartbeat number n = 1,2,. . .N−1. SD1 represents the short-term RR interval variability, and is known to be related to the HRV parameter in the time domain, SDSD [1]. As the long-term variability of the RR-interval, SD2 can be obtained in the PPA and it can also be expressed by the combination of SDSD and SDNN, which is another HRV parameter in the time domain. Some studies have considered that the parameters of SD1 and SD2 can be used as prognostic indices of certain diseases or as indices that reflect the autonomic state. In patients with chronic renal failure, it has been reported that the PPA parameters are smaller than those in healthy subjects [7]. Studies have shown that endurance exercise causes the PPA parameters to be increased by the progressive activation of the parasympathetic nervous system, so the widths of the Poincare plot can be used as replacement indices for linear HRV parameters to assess changes in that system [9]. Moreover, PPA has been applied to research regarding patients with depressed left ventricular function after an acute myocardial infarction, diabetic subjects, and patients with dilated cardiomyopathy [7].

2.2 Approximate and sample entropy

Approximate entropy (ApEn) eq.2.3, a parameter which indicates the rate of entropy, measures the likelihood through incremental comparisons between data patterns of a certain length m and the remaining data [5]. The lower values of

ApEn indicate a higher degree of regularity in the time series. ApEn is calculated as follows:

1. From *m* vectors *X (1) to X (N-m+1)* defined by

$$X(i) = [u(i), u(i+1), \ldots u(i+m+1)], 1 \le i \le N-m+1$$

…2.3

2. Define the distance $d[X(i), X(j)]$ between the vectors $X(i)$ and $X(j)$ as the maximum absolute difference their respective scalar components :

$$d[X(i), X(j)] = \max_{1,2,\ldots m}(u(i+k) - u(j+k))$$

…2.4

3. Define $v^m(i)$, such that $d[X(i), X(j)] \le r$ from which we define

$$c_i^m(r) = \frac{v^m(i)}{N-m+1}$$

…2.5

4. Take the natural logarithm of each $c_i^m(r)$ and average it over *i*

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_1^{N-m+1} \ln(c_i^m(r))$$

…2.6

5. Increase the *m to m+1* and repeat the steps 1 to 4 and find

$$\Phi^{m+1}(r) = \frac{1}{N-m} \sum_1^{N-m} \ln(c_i^m(r))$$

…2.7

6. Finally calculate ApEn for the data of length of N

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r)$$

…2.8

Sample entropy (SampEn) eq.2.8, is a parameter developed to reduce the bias by introducing self-matching count, to compensate for the shortcoming of ApEn [9]. This requires a simpler algorithm than ApEn and is significantly independent of data length [6]. SampEn can be obtained as follows:

1. For *m* vectors $X(1) to X(N-m+1)$ defined by,

$$X(i) = [u(i), u(i+1), \ldots u(i+m+1)], 1 \le i \le N-m+1$$

…2.9

2. Define the distance $d[X(i), X(j)]$

$$d[X(i), X(j)] = \max_{1,2,\ldots m}(u(i+k) - u(j+k))$$

…2.10

3. Define $v^m(i) = $ no of $X(i)$ such that
   $d[X(i), X(j)] \le r$ and $i \ne j$ from which we
   define $B_i^m(r) = \frac{1}{N-m+1} v^m(i)$ Here
   $i = 1, 2 \ldots N-m$ ..
   …2.11

4. Now Define $B^m(r) = \frac{1}{N-m} \sum_1^{N-m} B_i^m(r)$ and

$$A^m(r) = \frac{1}{N-m} \sum_1^{N-m} A_i^m(r) \qquad \ldots$$

2.12

5. Finally SampEn for a finite length of N can be taken as

$$SampEn(m, r, N) = -\ln\left(\frac{A^{m(r)}}{B^{m(r)}}\right)$$

…2.13

3. RECONSTRUCTION USING INTERPOLATION

Reconstruction is a process of inserting the predicted value in the actual missed dataset. Reconstruction has been performed by using different interpolation techniques. Interpolation is the way to constructing new data points within the range of the series [4]. The nearest neighbour, linear, cubic spline, and piecewise cubic Hermite (PCH) are the different interpolation methods and these interpolation processes has different algorithms of reconstruction. The difference in these interpolation methods while using the interp1 is shown in the fig2 and fig3. As it is evident that the higher order derivative for the methods becomes zero while using Nearest neighbor, Spline, PCH respectively. Therefore the difference between these interpolation techniques can only be seen in graphs. In this study, the interpolation used is Linear, zero interpolation and delete interpolation.

Linear interpolation technique is used with the inbuilt function (interp1). This interpolation technique comes in the category 1D interpolation. For the large value of data set the interpolation results are better as compared to the small valued dataset. [8]. Zero interpolation reconstructs the missed data value by taking the mean of the previous three values of the respective missed point. Delete interpolation practically does not incorporate any reconstruction as it only synthesis the missed data as it is. Suppose if there is an actual data of 1000 data points and if five points are randomly missed then delete interpolation will synthesis the remaining 995 points and gives the corresponding values where as in the other two interpolation techniques the 995 points are reconstructed back to 1000.
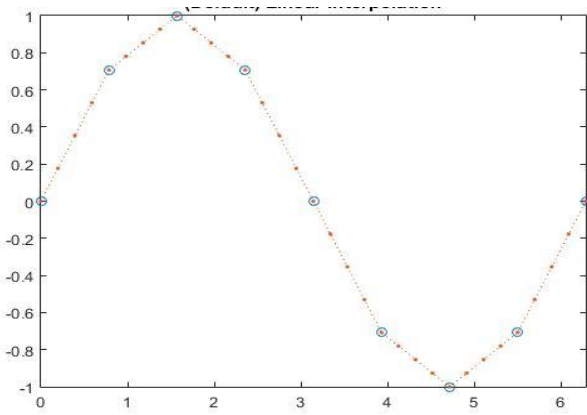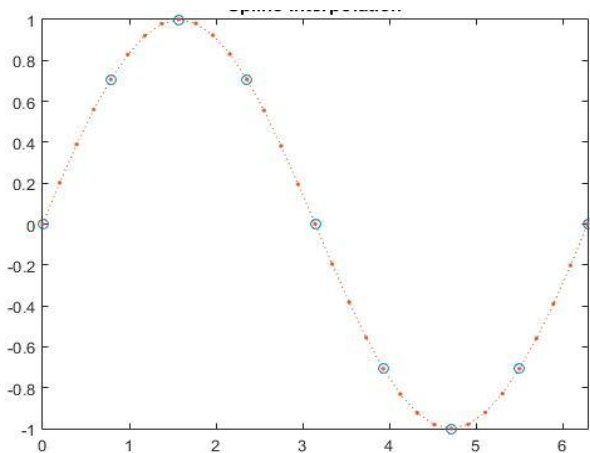
Fig 1: Showing Linear Interpolation

Fig 2: Showing Spline Interpolation

4. DATA AND METHODOLOGY

Data is obtained from www.Physionet.org of the 10 subjects of MIT-BIH Arrhythmia database. The data should be artefact free so as to obtain the smooth variations in the results. The artefact free data was processed in which random samples were removed with the frequency of 5 samples up to 50 samples and on those missed values the interpolation is done. To get the data with some missed values, the data has been randomly missed with the frequency of 5 data points up to 50 data points in the original data of 1000 samples and then with different interpolation the data is reconstructed back and compared. The nonlinear parameters were calculated for all the data sets and all the missed and interpolated values. After taking the averages the absolute relative error w.r.t. actual data was taken and graphs were plotted.
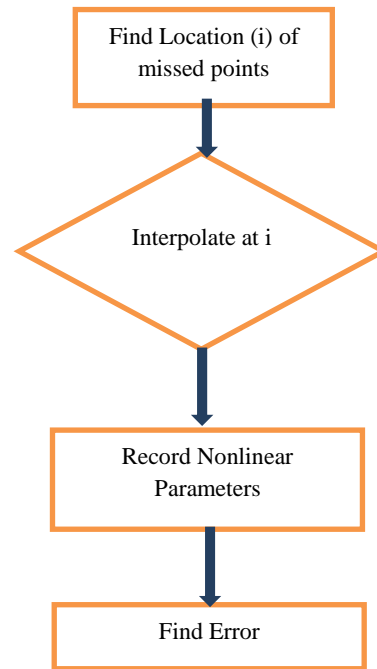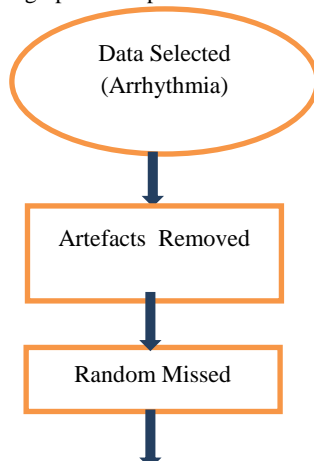

Fig 3: Depicting Methodology

5. RESULTS AND DISCUSSION

ApEn is known as the entropy parameter that is affected by the sample number and that includes the errors found between the estimation and theoretically predicted values. SampEn is therefore recommended when the sample number is small because of the independence on the number. Since the number of RR-interval data used in the simulation study does not exceed about 1000 samples, ApEn is considered to include some errors. These errors are considered to affect the errors from the missing data, and it is observed that ApEn is much more robust than SampEn in all the cases of interpolation.

In the actual missing data results, as shown in Fig. 5, the variation patterns of ApEn and SampEn differ from each other. Though the MATLAB functions for these were verified through additional tests on the different data sets, precise entropy values were returned for each parameter. It is therefore caused by the fact that the random nature of deletion incorporated in the process.

The values of SD1 and SD2 obtained are almost similar as the values are coming in the specific range but for different interpolation techniques the error is coming out to be different that means that if in case of SD1 the maximum error occurring is in the linear interpolation by using the interp1 function and minimum error is in the zero interpolation then for the same conditions and values the maximum and minimum error in the SD2 is occurring for the zero interpolation and delete interpolation respectively.

6. CONCLUSION AND FUTURE SCOPE

For the parameters SD1, SD2, Apen and Sampen the results are shown in fig.4, 5, 6, and 7. The Missed RR-Intervals has been reconstructed using various interpolation techniques. The frequency of missed intervals increases with a factor of 5 samples up to 50 samples in the dataset of 1000 samples that means insertion of data with interpolation is increasing with the factor of 5 samples in the final reconstructed data except delete interpolation. The robust technique that has been found for reconstructing a dataset which has random missed points is

delete interpolation. This is also concluded that for the three interpolation techniques discussed less error is present as compared to original data if the data is used as it is without reconstruction. This has been proved in the below figures 4,5,6,7 as when incorporating no interpolation techniques the absolute relative error is coming less for the average of 10 subjects in Poincare and entropy values as compared to the other interpolation techniques used namely linear and zero interpolation with one exception for SD1. Future scope of this work is very wide as during clinical usage by physicians the corrupt data can be avoided if there can be made a GUI showing non-linear parameters which can compare the original RR-Interval data with the new data. Also the analysis can be done for the grouped missed values in the rr-interval.

7. REFERENCES

[1] K. K. Kima, H. J. Baek, Y. G. Limb, K. S. Park, "Effect of missing RR-interval data on nonlinear heart rate variability analysis," *Comput Methods Programs Biomed Elsevier journals*, 2012

[2] P. Guerrero, C. Mailhes, and F. Castanié, "Lost Sample Recovering of ECG Signals in e-Health Applications," *Engineering in Medicine and Biology Society, 29th Annual International Conference of the IEEE*,pp. 31 - 34, 2007

[3] M. Brennan, M. Palaniswami, and P. Kamen, "Do Existing Measures of Poincare Plot Geometry Reflect Nonlinear Features of Heart Rate Variability," *Engineering in Medicine and Biology Society, Annual International Conference of the IEEE*,pp. 1342 - 1347, 2001

[4] G. Ganeshapillai, J. F. Liu, and J. Guttag, "Reconstruction of ECG Signals in The Presence of Corruption," *Engineering in Medicine and Biology Society, Annual International Conference of the IEEE*,pp. 3764 - 3767, 2011

[5] S. Pincus, Approximate entropy (Apen) as a complexity measure, CHAOS 5 (1995).pp. 110–117.

[6] M. Aboy, D. Cuesta-Frau, D. Austin and P. Micó-Tormos, "Characterization of Sample Entropy in the Context of Biomedical Signal Analysis," *Engineering in Medicine and Biology Society, Annual International Conference of the IEEE*,pp. 5942 - 5945, 2007

[7] C. Lerma, O. Infante, H. Perez-Grovas and M. V. Jose, "Poincare plot indexes of heart rate variability capture dynamic adaptations after haemodialysis in chronic renal failure patients," *Clin Physiol Funct Imaging*,pp. 72-80, 2003

[8] B. Efron, Missing data, imputation, and the bootstrap, J. Am. Stat. Assoc. 89 (1994) 463–475

[9] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol.-Heart C 278 (2000) H2039–H2049

8. APPENDIX

Fig 4: Showing Error in SD1 for different interpolation techniques

Fig 5: Showing Error in SD2 for different interpolation technique

Fig 6: Showing Error in ApEn for different interpolation technique

Fig 7: Showing Error in SampEn for different interpolation technique