A Study on the Quantitative Regularity Measures That Are Suitable for Biological Signal Analysis - Standard Data and 24 Hour R-R interval Analysis

°남 영 한, 이 종 민, 한 주 만, *박 광 석
서울대학교 공과대학 협동과정 의용생체공학 전공, *의과대학 의공학교실
°Y. H. Nam, J. M. Lee, J. M. Han, *K. S. Park
Interdisciplinary Program in Medical and Biological Engineering Major,
*Department of Biomedical Engineering, College of Medicine,
Seoul National University

ABSTRACT - We tested the capability of Pointwise Correlation Dimension(PD2), Approximate Entropy (ApEn) and LZ complexity, as alternative measures of a biological signal. For this purpose, we analyzed standard data and a healthy child's 24-hour heart rate variability. Our conclusion is as follows. First, PD2, ApEn and LZ complexity can be used for discerning chaotic attractor, white noise, and periodic signal. Second, these measures show different characteristics on day and night. Third, these measures can be used for detecting time-varying characteristics of biological signals.

INTRODUCTION

It is widely known that heart rate variability has irregular component.[1] To quantify a regularity of a time series, various complexity measures are used, and Correlation dimension(D2) is popular measures to quantify regularity. However, D2 requires very long calculation time, and assumes the time series to be stationary. D2 cannot identify time- varying complexity; that is, D2 cannot be utilized for real-time processing of biological signal. For these shortcoming of D2, alternative measures, which are PD2, ApEn, and LZ complexity, are recently used.

Correlation Dimension(D2) is calculated from correlation integral, which computes the density of recurrence plot,

$$C^m(r) = \frac{\text{number of times} \mid \mathbf{D}^m(i) - \mathbf{D}^m(j) \mid \le r}{N(N-1)}$$

where m is an embedding dimension, $\mathbf{D}^m(i)$ is an i 'th embedded vector and N is number of embedded vectors. D2 is defined as a slope of log r vs. log $C^m(r)$ plot.

Pointwise-D2 (PD2)[2] does not utilize whole time series. PD2 calculate D2 at i 'th point, over only predefined number of time series near i 'th point. As doing so, stationary presumption can be avoided, fast calculation is available, and detecting time varying characteristics of given time series is possible.

ApEn[3] measures an inverse of average logarithm of conditional probability that

$$d[\mathbf{D}^m(i), \mathbf{D}^m(j)] \le r$$

given that

$$d[\mathbf{D}^{m+1}(i), \mathbf{D}^{m+1}(j)] \le r$$

where $d[\mathbf{D}^m(i), \mathbf{D}^m(j)]$ is the maximum of differences of each embedded vector components.

Therefore, ApEn has a higher value when embedding dimension is smaller than or near of an attractor dimension, and has a lower value when embedding dimension is higher than an attractor dimension.

LZ complexity[4], γ is an algorithmic complexity for sequences of finite length. γ is related with the order that is retained in the sequence.

c(s) is a relative index to quantify the regularity of *given* sequences by reflecting the order retained in the sequence.

The complexity measure b(n) for random sequences is defined as follows.

$$b(n) = \frac{hn}{\log_k n}$$

where h= the normalized source entropy

$$= \sum_{i=1}^{k} p_i \log_k p_i$$

k= the number of elements in the alphabet

Then, the normalized complexity is

$$\gamma = \lim_{n \to \infty} \frac{c(s)}{b(n)}$$

METHODS & RESULT

1. Standard time series analysis

We generated the standard time series that is composed of 10,000 points (Table 1).

We selected parameters of each regularity measures so that they have different value at noise, periodic,

low-dimensional chaotic range.

For ApEn, m=2, window size N=1000 will produce reasonable statistical validity.[3] ApEn could discriminate random sequence and low-dimensional chaos when ApEn of random sequence is higher value. The ApEn of random sequence increases with decreasing r. On the other hand, too small r value will produce very small ApEn. Hence, optimization of r is required. In this paper, we propose r to be (standard deviation of time series) / 5.

As LZ complexity is the ratio of regularity of random series and given time series, LZ complexity of random series should be 1 theoretically. For LZ complexity of random series to be 1, window size should be large or encoding level be small enough. However, increasing window size will require more calculation time, decreasing encoding level will result in worse resolution, therefore, trade-off is required. In this paper, we selected window size to be 1000 as other algorithm, encoding level to be 10.

For PD2, the parameter that we can select is only window size, in this paper, we selected window size to be 1000 as other algorithm.

For reducing calculation time, we use skipping of window, skipping value are 200 in each algorithm.

According to above determination of parameters, we got the resulting standard data analysis plot as shown in Fig. 1.

seq. #	1~2000	~4000	~6000	~8000	~10000
kind of sequence	White Noise	sinusoid	Lorenz Map	Henon Map	Quadratic Map

Table 1 Standard time series

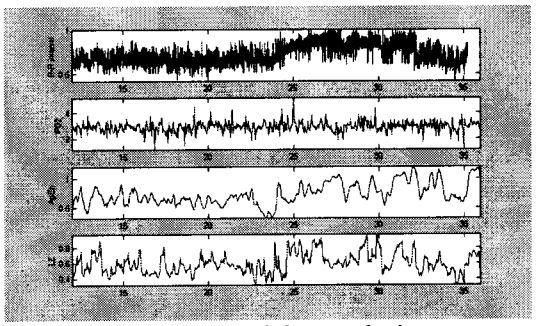

Fig. 1. Stand data analysis

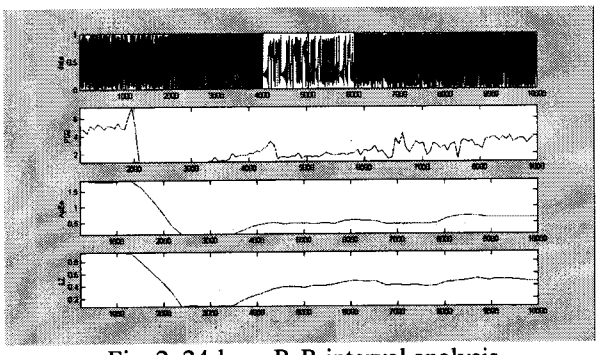

Fig. 2. 24-hour R-R interval analysis

2. 24 hour R-R interval analysis

We analyzed a healthy child's 24-hour R-R interval using three regularity measures whose parameters are determined by 1. of this section. The result is shown in Fig. 2.

Our conclusions of Fig. 2. are as follows.

First, the each complexity values of R-R interval are in the chaotic range that is determined by 1. of this section

Second, complexity values are higher in night than in day, this result corresponds with spectrum analysis.

DISCUSSION

The fluctuation is larger in PD2 than in two other measures when using the same window size and skip.

In the analysis of R-R interval, LZ complexity and ApEn show similar patterns. In particular, LZ complexity and ApEn of night range R-R interval could be used for sleep stage analysis.

We will analyze more children's R-R interval data using these method, find clinically appropriate conclusion .

CONCLUSION

We determined parameters that are suitable for analyzing biological signal for the three measures; Pointwise Correlation Dimension, Approximate Entropy, LZ complexity. We concluded that these three measures can be used for real-time characteristics analysis of biological signal.

REFERENCES

[1] Kanters JK, Hojgaard MV, Agner E, Holstein-Rathlou NH, "Short- and long-term variations in non-linear dynamics of heart rate variability", Cardiovasc. Res., vol. 31, pp. 400-409, March 1996

[2] Skinner JE, Carpeggiani C, Landisman CE, Fulton KW, "Correlation dimension of heartbeat intervals is reduced in conscious pigs by myocardial ischemia", Circ Res 4/68:966-276. 1991

[3] Steven M. Pincus, "Approximate entropy as a measure of system complexity", Proc. Natl. Acad. Sci. USA, vol. 88, pp. 2297-2301, March 1991

[4] Abraham Lempel, Jacob Ziv, "On the complexity of finite sequences", IEEE tr. on information theory, vol rr-22, no.1, pp75-81, January 1976