A Nonlinear Model for Predicting ECG R-R Interval Variation Based on the Evolutionary Computation Approach

Bo Yeon Kim[1], Yun Seok Chang[2], and Kwang Suk Park[3]

[1] Department of Electrical and Computer Engineering, Kangwon National University Chuncheon, Korea
[2] Department of Computer Engineering, Daejin University, Pocheon, Korea
[3] Department of Biomedical Engineering, College of Medicine, Seoul National University, Seoul, Korea

Abstract. The study proposes a modeling method for nonlinear system, which predicts characteristics of the ECG R-R interval variation. For determining model equation, we adopted genetic programming method that the chromosome represented the model equation consisting of time-delayed variables, constants, and four arithmetic operators. By the genetic programming the regressive nonlinear equations were produced and evolved to find the optimal model equation which could simulate the spectral, statistical and nonlinear behavior of the given R-R interval dynamics. Experimental results showed that the evolutionary approach could find the equation that simulates the spectral and chaotic dynamics of the given signal. Therefore the proposed evolutionary approach is useful for the system identification of the nonlinear biological system.

1 Introduction

Researchers have often considered that a biological system produces signals being regular and periodic under the normal state. Recently, some clinical observations have indicated that a biological signal seems to be irregular and non-periodic for the normal states, but surprisingly becomes periodic and regular for the abnormal state, for instance, EEG of petitmal epilepsy, heart rate.[12,13,7] Therefore the chaotic biological signals can be represented by an equation or a group of equations. In our study, we have taken into consideration a deterministic approach for describing a nonlinear biological system and its intrinsic mechanisms. We had attempted to find a nonlinear equation with some control parameters to describe the chaotic characteristics of a biological system.

Traditionally , AR model is the most common method to fit a signal from a linear system. In PAR, a current heart rate HR(k) is defined by a quadratic nonlinear function of sum of former heart rate products products

$$\sum_i \sum_j b_{ij} x(k-i)x(k-j)$$

P.M.A. Sloot et al. (Eds.): ICCS 2003, LNCS 2657, pp. 378–386, 2003.

plus the AR component. In bilinear, a current heart rate $x(k)$ is defined by a quadratic nonlinear function of sum of former heart rate products

$$\sum_i \sum_j b_{ij} x(k-i) e(k-j)$$

plus the ARMA component, where $e(k-j) = x(k-j) - x(k-j)$, b_{ij} is the coefficients to be found and $e(j)$ is the error that the difference of the measured vaue of x from that estimated at the j-th time delay. PAR requires long computer running time (input signals normally less than 250 points) and require the predetermined structure of the function for fitting.

Unlike the previous PAR and Bilinear, in the study adopted Genetic programming (GP) approach that was not based on AR method. GP does not require the predetermined structure of the function for fitting and finds the structure of the function as well as the values of the coefficients which represent best given input signals. Genetic programming (GP) is known one of the searching methods, which evolves solutions iteratively until, by natural selection, it finds the best satisfied solution to given criteria. The study accounted for the R-R intervals of ECG from a normal child which were shown to be irregular and non-periodic. Aim of the study was to determine a nonlinear equation optimally fitted to the characteristics of the R-R intervals using the genetic programming.

2 Method

2.1 Proposed Genetic Programming Method for Modeling a Chaotic Signal

The developed method in this study employs a genetic programming (GP) approach and finds a polynomial function consisting of the constants and variables and optimally fitting to a measured physiological signals. The polynomial function that the proposed method predicts is composed of arithmetic operators, time-delayed variables and constants and has the form defined by

$$X_n = f\left(a_1, a_2, ..., a_m : X_{n-1}, ..., X_{n-k}\right) \tag{1}$$

where $a_1, ..., a_m$ are constants and $X_{n-1}, ..., X_{n-k}$ are time-delayed variables.

If a chromosome is defined as a polynomial expression, GP creates many shapes of polynomials by the crossover and mutation operation. Since GP is able to handle polynomials, it can be appropriate to find a polynomial in the shape of the equation (1) which characterizes or models a given signal.

To find the best model equation for a given time series using GP, we took the following steps.

1st step : design of chromosomes
2nd step : determine fitness function
3rd step : set the genetic operators and execution control parameters
4th step : run the program for an optimal polynomial

Details of each step will be given below.

2.2 Design of Chromosomes

Table 1 shows the elements and the structure of the chromosomes for finding the model equation of ECG R-R intervals. Since the number of variables was not known for the measured signal that is ECG R-R intervals, we set it here as large as the computing time and accuracy reasonably acceptable. In the present study we took 9 variables for the ECG RR interval signals. The terminal of a chromosome was a set of constants and time-delayed variables. The value of the constants was ranged from 0.01 to 9.99. The primitive function was a set of arithmetic operators which were used to construct a polynomial. The operators were addition, subtraction, multiplication and division, and in this study it was assumed that a nonlinear equation could be approximated using those four operators.

Since each model equation has the same left-hand part as shown in table 1, the left-hand term and equal operator, '$X_n =$', was excluded to set a chromosome. Therefore, a chromosome represents only the right-hand part of the model equation, and is constructed by time-delayed variables, constants, and arithmetic operators.

Table 1. Structure of chromosomes for the nonlinear systems. Where $X_{n-1}, ...X_{n-k}$ are time-delayed variables and $a, b, c, ..., k$ are constants as control parameters

	R-R interval variation
Input data	Time series of normal R-R interval
Model equation	$X_n = f(a, b, c, ..., k; X_{n-1}, ..., X_{n-9})$
Chromosome	$f(a, b, c, ..., k; X_{n-1}, ..., X_{n-9})$ in prefix notation
Terminals	$\{a, b, c, ..., k; X_{n-1}, ..., X_{n-9}\}$
Number of time-delayed variable	9
Number of constants	no restriction
Primitive functions	{+,-,x,/}

2.3 Determine Fitness Function

The measure of the fitness was defined by the inverse of the weighted sum of the parametric differences between the original and the estimated time series. The parametric differences are the differences on the 6 parameters (mean, variances, regional histograms, range, fractal dimensions of projected dots, box-counting dimension of signal, correlation integrals and square sum of difference of the two signals.[1,2,4,5,15,9,14]

The regional histogram means the number of points in the equally segmented regional bin from the minimum value to maximum value. Box-counting dimension of signal was computed as following function.

$$D_0 = \lim_{x \to 0} \frac{\ln M(\epsilon)}{\ln(1/\epsilon)}$$

where ϵ is the length of 2-dimensional box, $M(\epsilon)$ is the minimum number of boxes to cover the signal.

2.4 Genetic Parameters for the Execution Control

The proper rates for three genetic operations were adjusted in experiments. For each operation, 1 or 2 chromosomes were selected by the roulette wheel method as parent chromosomes, with a probability based on the fitness.

In the present study, we set the range of the values of the control parameters related to the execution of GP as following :

- the number of chromosomes in a generation: 100 to 500
- the number of generations : 50 to 300
- crossover rate : 0.3 to 0.7
- mutation rate : 0.1 to 0.5
- reproduction rate : 0.1 to 0.3

Note that the chromosomes produced by the GP are all the candidates for a final polynomial. The total number of the candidates are created under the condition are 5000 to 150000.

If the iteration number of GP is reached to the maximum, the present GP is terminated. The model function best-fitted for the input signal can be constructed from the chromosome which has the maximum value of the fitness measure in all generations.

2.5 Simulation Procedure for an Optimal Polynomial

In the computer program for simulation, the chromosome is expressed by the prefix notation where the operator precedes the operands and is manipulated by a tree structure. The simulation program runs the following 3 steps :

1. Generate an initial population with random proportions of the functions and terminals defined in Table 1. The chromosomes produced by the present GP are the right-hand part of the model equation.
2. Iteratively perform the following substeps until the terminal criterion has been satisfied.
   a) Evaluate each chromosome in the population and assign it a fitness value. From each chromosome, generate time series using the same initial value of the given time series and calculate the statistical and nonlinear characteristics for generated time series. Measure the fitness using calculated values of the statistical and nonlinear characteristics.
   b) Create population for the new generation by applying the three genetic operations as follows. Those operations are applied to chromosomes in the population with probabilities based on their fitness. During genetic operations, the reselection of chromosomes is allowed.

   i. Reproduce an existing chromosome in a population of the current generation by copying it into a population of the next generation.
   ii. Create two new chromosomes from two parent chromosomes by crossover using 2-point crossover, that is, the middle part of 3 partitions is switched.
   iii. Create a new chromosome by mutating a randomly chosen part with randomly produced polynomial.
3. Choose a chromosome that has the maximum fitness. And then, to construct a model equation, convert the best-fitting chromosome from the prefix notation to the in-order notation and added '$X_n =$' in front of the converted expression.

3 Experiments and Results

ECG was recorded from 5 normal child relaxed for 15 minutes and was sampled at 1000Hz and digitized via 12 bit A/D converter, and then stored on a PC at Seoul National Children's Hospital. Application of the R wave detection algorithm suggested by Thomkins et al to the acquired signal resulted in the R-R interval time series was obtained to be used as an input signal to the present genetic program.

The genetic programming for modeling a nonlinear behavior of the biological signals was written in C compiler and run on the SUN SPARC 20. The running time varied by the values of the execution control parameters and the shapes of the fitness function. It was found that, as expected, the simulated results became more accurate as the number of generations and chromosomes increased. The optimum values of the execution control parameters which enabled the program to give the most accurate results for minimum time were as follows;

- the number of input data points: 512
- the number of chromosomes in a generation: 500
- the number of generations : 300
- crossover rate : 0.3
- mutation rate : 0.49
- reproduction rate : 0.21

Figure 1 shows the measured R-R interval time series, where the label n indicates the $n-th$ heart beat and $X[n]$ is the interval between beats in units of millisecond. The phase-space diagram of the signal is displayed in figure 2 and shows that the R-R intervals varies in the boundary so that it is chaotic. This R-R interval signal was fed into the proposed GP with the values of the control parameters and the computed results are given in figure 3. The best chromosome with the maximum fitness is

$$-X_{n-9} \times -X_{n-8}X_{n-1} - +X_{n-9} \times 0.07X_{n-8} \times 0.07X_{n-9}$$

and the responding equation of the chromosome is

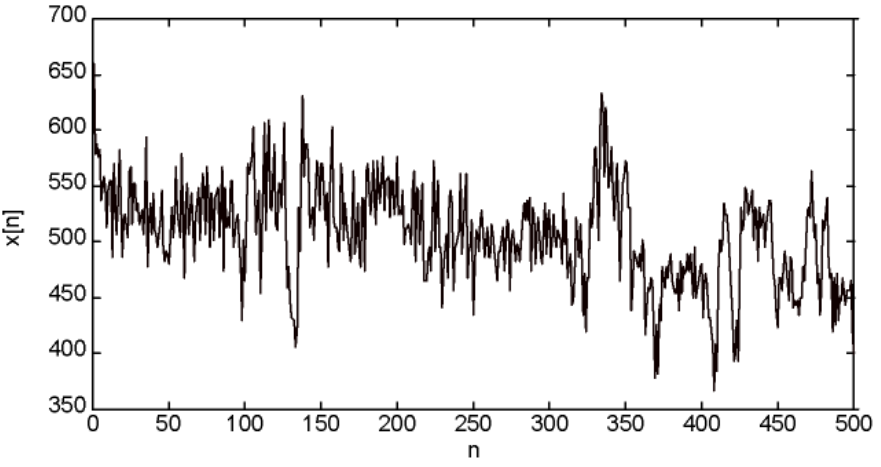

Fig. 1. Time series of R-R intervals of ECG; input time series

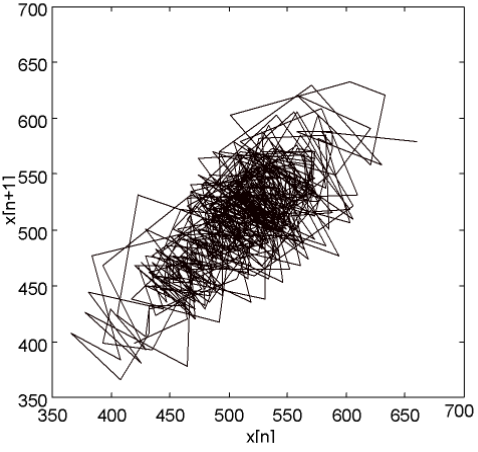

Fig. 2. Phase space diagram of time series in figure 1

$$X_n = X_{n-9} - (9.93X_{n-9} - 0.07X_{n-8})(X_{n-8} - X_{n-1})$$

Note that, although we used nine time-delayed variables for a chromosome, only 3 time-delayed variables, $X_{n-1}, X_{n-8}, X_{n-9}$ appeared in the computed equation. To show up the chaotic property of the predicted signal, its strange attractor is depicted in figure 4. The statistical and chaotic characteristics for the original and predicted times series of ECG R-R interval are summed up in table 2. It shows that the chaotic properties are very accurately predicted, while there are some insignificant differences in a statistical parameter such as mean.

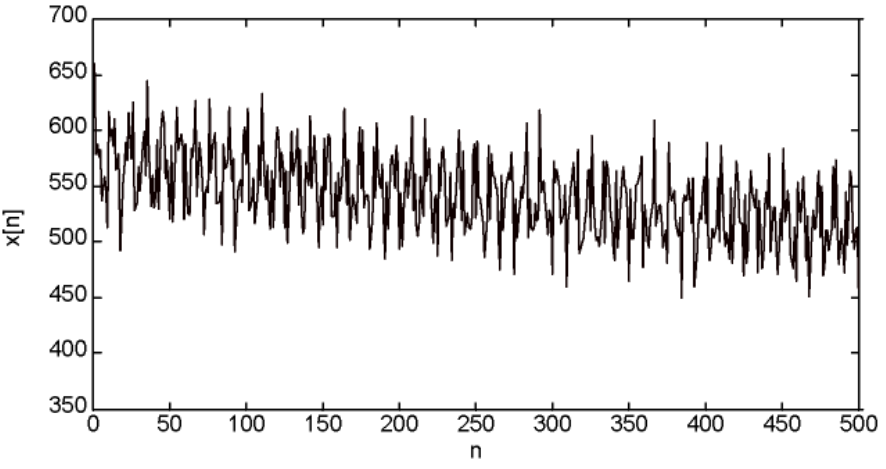

Fig. 3. Modeling results for the time series of R-R interval; time series generated by the estimates equation $X_n = X_{n-9} - (9.93X_{n-9} - 0.07X_{n-8})(X_{n-8} - X_{n-1})$ with the same initial values as figure 1

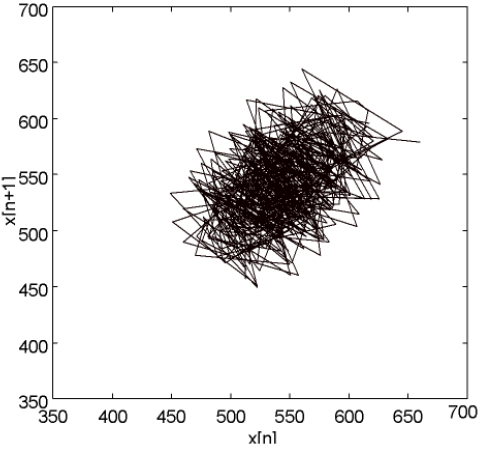

Fig. 4. Phase space diagram of time series in figure 3

Other model equations in table 3 showed similar nonlinear characteristics, fractal dimension, even though estimated model functions vary.

4 Discussion

The evolutionary approach has been proposed for finding a nonlinear function optimally fitting for the given chaotic time series. In the case of modeling ECG R-

Table 2. Statistical and nonlinear characteristics of the original and estimated times series of normal ECG R-R intervals

	Original	Estimated
Mean	505.92	541.18
Variance	2.20	1.33
Range	$366 \sim 660$	$449 \sim 660$
Projected fractal dimension	0.752	0.772
Fractal dimension	1.09954	1.09824

Table 3. Model functions and their fractal dimension

	Model equation	Fractal dimension	Error
Original		1.328	
Result 1	$X_n = X_{n-5} + \frac{X_{n-5}^2(X_{n-3}-X_{n-2})}{1.56X_{n-3}X_{n-5}X_{n-6}}$	1.368	3.0%
Result2	$X_n = X_{n-9} + (0.93X_{n-9} - 0.07X_{n-8})(X_{n-1} - X_{n-8})$	1.373	3.4%
Results3	$X_n = X_{n-9} + (X_{n-1} - X_{n-8})/(5.42X_{n-1})$	1.405	5.8%

R intervals, the signal generated by the equation represented by the chromosome were seen to be not similar to the input signal in time domain, while its spectral, statistical and chaotic characteristics were shown to be reasonably similar as those of the input signal.

To find the model function we optimize the control parameters. Unlike the conventional use of Genetic Programming, the mutation rate for modeling ECG R-R intervals was required to be larger than the crossover rate. The value of the mutation rate is also required to be below 0.5, so that the model equations may predict the nonlinear characteristics of input signals better. This may indicate that the mutation rate plays an important role in searching the huge solution space and in escaping the local minimum. It was found that the structure of the fitness function was closely associated with the degree of complexity or nonlinearity of the signals, that is, the number of the independent parameters to construct the fitness functions increased with the complexity of the signals.

In general, the strange attractors were shown to be quite similar between the predicted and input signals. Figure 4 shows that the model equation appropriately represented the statistical and chaotic characteristics of the source signals, even if the time history of the predicted signal was not fitted well. The proposed approach was able to yield a nonlinear equation which could not be well coincide in time domain but precisely described its statistical and chaotic characteristics. Therefore the proposed evolutionary approach is useful for the system identification of the biological system. Further optimization technique in selecting more appropriate fitness function may be needed to obtain an improved time domain fitting and the authors leave it as a future study.

Acknowledgement. This work is the result of research activities of Advanced Biometric Research Center (ABRC) supported by KOSEF.

References

1. Albano, A. M., Muench, J., Schwartz, C., Mees, A. I., Rapp, P. E.: Singular-value decomposition and the Grassberger-Procaccia algorithm: Phys. **Rev. A. 38** (1988) 3017–3116
2. Broomhead, D. S., King, G. P.: Extracting qualitative dynamics from experimental data. **Physica D 20** (1986) 217–236
3. Gleik, J.: Chaos : Making a New Science: New York: Viking Penguin Inc. (1987)
4. Grassberger, P., Procaccia, I.: Measuring the Strangeness of Strange Attractors. **Physica D 9** (1983) 198–208
5. Gulick, D.:Encounters with Chaos: McGraw-Hill Inc. (1992)
6. Hamilton, Patrick S., Tomkins, Willis J.: Quantitive investigation of QRS detection rules using the MIT/BIH arrythmia database:IEEE Transactions on Biomedical Engineering. **BME-33** (1986) 413-422
7. Kaplan, D., Glass, L.: Direct test for determinism in a time series:Phys. **Rev. Lett. 68** (1992) 427–430
8. Koza, J. :Genetic Programming II: Automatic Discovery of Reusable Programs: The MIT Press (1994)
9. Liebovitch, L. S., Toth, T.: A Fast Algorithm to Determine Fractal Dimensions by Box Counting: Phys. **Lett. A 141** (1989) 386–390
10. Mead, W. C., Jones, R. D., Lee, Y. C., Barnes, C. W., Flake, G. W., Lee, L. A., O'Rourke, M. K.: Prediction of Chaotic Time Series using CNLS-Net-Example: The Mackey-Glass Equation: Proc. of the Workshop on Nonlinear Modeling and Forecasting. **12** (1990) 39–71
11. Mees, A. I.: Tessellation's and dynamical systems': Proc. of the Workshop on Nonlinear Modeling and Forecasting. **12** (1990) 3–24
12. Pool, R.: Is It Healthy to Be Chaotic: Science. **243** (1989) 604–607
13. Rigney, D. R., West B. J.: Chaos and Fractal in Human Physiology: Scientific American. (1990) 35–41
14. Sano, M., Sato, S., Sawada, Y.: Global Spectral Characterization of Chaotic Dynamics': Prog. Theor. Phys. **76** (1986) 945–948
15. Kim, B. Y., Park, K. S.: Modeling of nonlinear sequence based on the Modified Genetic Algorithm: Nonlinear Analysis. **36** (1999) 707–720
16. Davis, Lawrence: Handbook of Genetic Algorithms: Van Nostrand Reinhold. (1991)