

유전자 프로그래밍을 이용한 생체 신호의 비선형 특성 모델링에 관한 연구

김 보연, 박 광석*

서울대학교 협동과정 의용생체공학전공

*서울대학교 의과대학 의공학교실

A study on the Modeling of Nonlinear Properties of Biological Signal using Genetic Programming

Boyeon Kim, Kwangsuk Park*

Interdisciplinary Program in Medical and Biological Engineering Major,

Seoul National University

*Dept. of Biomedical Engineering, College of Medicine, Seoul National University

Abstract

Many researchers had considered biological systems as linear systems. In many cases of biological systems, the phenomena that show the regular and periodic dynamics are considered the normal state. However, some clinical experiments reported, in some cases, the periodic signals represented the abnormal state. We assume that signals from human body system are generated from deterministic, intrinsic mechanisms and can be represented a simple equation that show nonlinear dynamics dependent on control parameters. The objective of our study is to model a nonlinear dynamics correctly from the nonlinear time series using the genetic programming method; to find a simple equation of nonlinear dynamics using collected time series and its nonlinear characteristics. We applied genetic programming to model RR interval of ECG that shows chaotic phenomena. We used 4 statistic measures and 2 fractal measures to estimate fitness of each chromosome, and could obtain good solutions of which chaotic features are similar.

1. 서론

일반적으로, 생체 신호는 선형 시스템으로 인식되어왔다. 생체 시스템이 안정적이고 정상상태일 때 생체 신호는 주기적이고 예측 가능한 특성을 보인다. 그러나, 최근에는 정상상태일 때 비주기적이고, 비정상 상태인 경우에 주기적인 특성을 보이는 생체 신호들이 많이 관찰되고 있다. 예를 들면, 심실 빈맥인 경우의 심전도 신호와 소발작 간질 환자의 뇌파 신호등이 이러한 생체 신호에 해당된다.

본 연구에서는 비선형 특성을 보이는 생체 신호는 생체 시스템의 결정적이고 내부적인 기전에 의하여 생성되며, 이러한 비선형 신호는 logistic map 함수에서 제어 파라미터에 따라 비선형 특성이 나타나는 것처럼 제어 파라미터에 의존하는 방정식으로 표현 가능하다고 가정하였다. 따라서 비선형 특성을 보이는 생체 신호를 수집하여 유전자 프로그래밍을 이용하여 주어진 생체 신호와 유사한 비선형 특성을 가지는 방정식을 생성해내고자 한다. 유전자 프로그래밍 방법은 일종의 최적 탐색 기법으로 알려져 있으며, 임의의 해의 집합을 자연 선택 방법을 응용하여 반복적으로 진화 시킴으로써 주어진 조건을 만족하는 최적의 해를 찾아내는 방법이다. 자연 선택에 의한 진화의 방법이 다른 탐색 기법과 구별되며, 비록 처음에는 임의의 해의 집합으로부터 시작하지만 주어진 조건을 만족하는 해를 다음 세대로 유전시키는 것은 임의 탐색(random search) 방법과도 구별된다.

2. 방법

유전자 프로그래밍 방법을 생체 신호의 비선형 특성 모델링에 적용하기 위하여 다음을 단계별로 수행하였다.

- (1) 문제 및 해답의 표현 정의
- (2) 적합도 판정 함수 정의
- (3) 수행 제어 파라미터 정의
- (4) 유전자 프로그래밍 수행 종결 조건 및 해답 선택 방법 정의

문제 정의

비선형 동특성을 보이는 생체 신호는 제어 파라미터에 의하여 영향을 받으며 시간-지연 변수(time-delayed variables)로 이루어진 방정식으로 가정하였다. 따라서 유전자 프로그래밍으로 모델링한 방정식을 식(1)과 같이 정의하였다.

$$X_n = f(a, b, c, \dots; X_{n-1}, X_{n-2}, \dots, X_{n-k}) \quad \dots\dots(1)$$

여기서, a, b, c, ...는 제어 파라미터의 기능을 수행하는 상수(constant)이고, 변수 X_{n-1}, \dots, X_{n-k} 는 시간-지연 변수이다. 유전자 프로그래밍으로 생성할 식, 즉 유전자는 모두 식(1)과 같은 형식이므로 등호 및 변수 X_n 은 모든 식에 나타나기 때문에 생략하여, 유전자 프로그래밍으로 방정식은 등호의 오른쪽 부분인 연산식 $f(a, b, c, \dots; X_{n-1}, X_{n-2}, \dots, X_{n-k})$ 을 생성한다. 연산식은 상수와 시간 지연 변수, 그리고 사칙연산자(덧셈, 뺄셈, 곱셈 그리고 나눗셈)로 구성된다. 연산식을 구성하는 상수는 0.01부터 9.99까지의 실수로 소수 두 자리로 제한하였으며, 시간 지연 변수는 X_{n-1}, \dots, X_{n-9} 로 제한하였다. 변수의 중복, 연산자의 중복에는 제한이 없으며, 생성할 식의 길이에도 제한이 없다. 제어 파라미터 값은 유전자 프로그래밍으로는 소수 두 자리로 생성하지만 실수 연산항의 계산 결과 보다 정밀한 값이 구해질 수 있다.

유전자 프로그래밍으로 모델링한 방정식의 예를 들면 다음과 같다.

$$3.64X_{n-1}^2 + 1.32 - X_{n-1}$$

$$X_{n-1}^3 + (7.62/1.27)X_{n-1}^2 - 0.83X_{n-1} + 3.98 \times 0.37$$

위 식의 실수 연산항을 계산하면

$$X_{n-1}^3 + 6X_{n-1}^2 - 0.83X_{n-1} + 1.4726$$

이다.

위와 같이 생성한 식은 실제로는 다음과 같은

의미이다.

$$X_n = 3.64X_{n-1}^2 + 1.32 - X_{n-1}$$

$$X_n = X_{n-1}^3 + 6X_{n-1}^2 - 0.83X_{n-1} + 1.4726$$

유전자 프로그래밍 시뮬레이션 프로그램의 편의를 위하여 연산식을 연산자를 피연산자의 앞에 위치시키는 전치 표현(prefix representation)으로 나타낸다. 즉, 연산식 $a+b \times c$ 는 전치 표현으로는 $+a \times bc$ 이다. 전치 표현은 적합도 판정을 위하여 방정식으로부터 시계열 데이터의 생성시 계산이 편리한 이점이 있다.

적합도 판정 함수 정의

유전자, 즉 생성한 연산식의 적합도를 측정하기 위하여 연산식으로부터 시계열 데이터를 생성하여 주어진 생체 신호와 비교하였다. 비교에 사용된 특성으로는 통계적 특성과 혼돈 특성을 사용하였으며, 적합도(fitness)는 각 특성의 차이값의 총합의 역수로 정의하였다.

즉, 적합도 =

- 1 / (두 시계열 데이터의 차이
- + 평균의 차이
- + 분산의 차이
- + 구간별 히스토그램의 차이
- + 값의 범위의 차이
- + fractal dimension의 차이
- + correlation dimension의 차이)

여기서, 첫 번째 측정치는 두 신호의 유사성 정도를 측정하는 것이고, 다음 네개의 측정치는 두 신호의 통계적 특성, 그리고 마지막 두개의 측정치는 비선형 특성의 유사한 정도를 측정하는 값이다. 적합도를 구성하는 각 항은 가중치를 곱하여 실험하였다.

수행 제어 파라미터의 결정

유전자 프로그래밍을 제어하는 몇 가지 파라미터 값을 정한다.

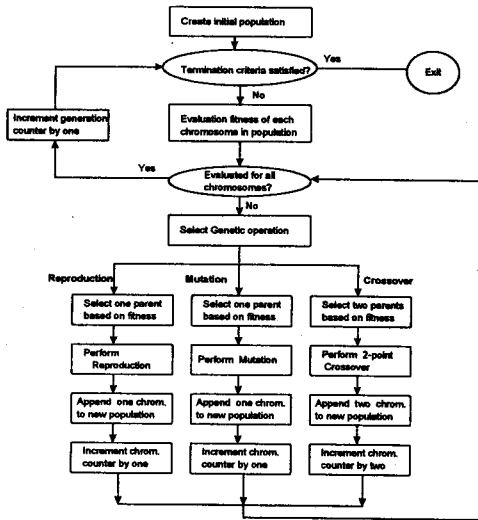
수행에 관련된 파라미터는 다음과 같다.

- 한 세대에서의 최대 유전자의 수: 200 ~ 300
- 최대 세대의 수 : 100 ~ 300
- crossover rate : 0.5 to 0.7
- mutation rate : 0.1 to 0.3
- reproduction rate : 0.1 to 0.15

수행 종결 조건 및 해당 지정 방법

세대 교체가 최대 반복 세대 수에 도달하면 수행을 종결한다. 수행 중에 생성된 모든 유전자 중에서 적합도가 최대인 유전자를 해당, 즉 주어진 생체 신호를 모델링한 방정식으로 정한다.

<그림 1> 유전자 프로그래밍의 수행 절차



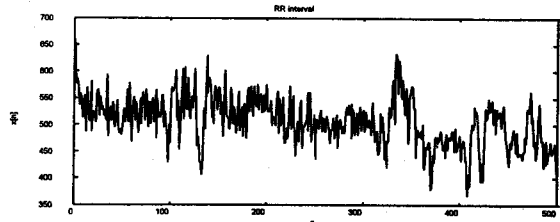
3. 실험 결과

심전도 신호는 정상인이 편안히 누운 상태에서 15분정도(900초) 동안의 데이터를 A/D 변환기를 통하여 600Hz로 샘플링하여 수집하였다. 수집된 심전도에서 R-R 간격을 측정하여 실험에 사용하였다. 그림 2와 같이 하여 시계열 데이터, phase-space diagram, lyapunov exponent를 구한 결과 비선형적임을 알 수 있었다.

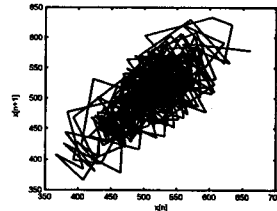
유전자 프로그래밍의 실행 제어 파라미터는 표1에 보인다. 그림 2의 시계열 데이터를 유전자 프로그래밍에 입력하여 모델링하면 $+/- X_{n-3}X_{n-2}1.56// * X_{n-5} / * X_{n-3}X_{n-6}X_{n-8}X_{n-5}X_{n-5}$ 으로 이를 방정식으로 완성하면 $X_n = X_{n-5} + X_{n-8}^2 (X_{n-3} - X_{n-2}) / X_{n-3}X_{n-6}$ 이다. 위 방정식에 생체 신호와 동일한 초기값을 사용하여 1000포인트 시계열 데이터를 생성하면 그림 3과 같다.

<표 1> 실행 제어 파라미터

최대 반복 세대 수	100
한 세대내의 유전자 수	200
Crossover rate	0.7
Mutation rate	0.21
Reproduction rate	0.09

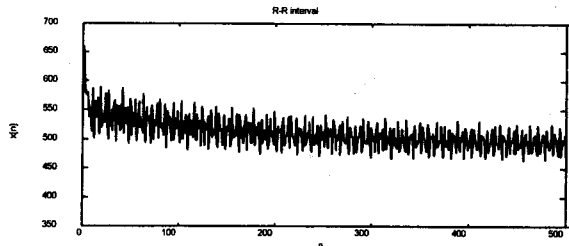


(a) 시계열 데이터

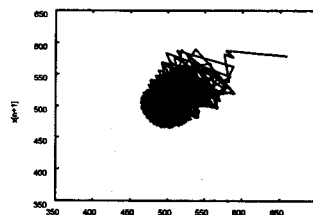


(b) phase-space diagram

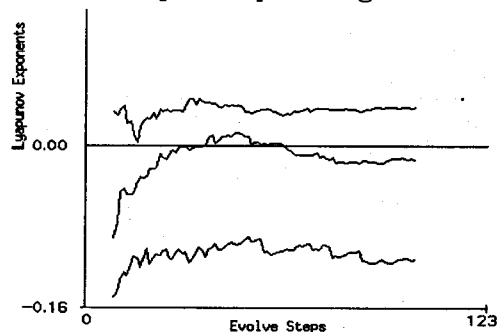
<그림 2> 심전도 RR 간격의 시계열 데이터



(a) 시계열 데이터



(b) phase-space diagram



(c) Lyapunov exponent

$$X_n = X_{n-5} + X_{n-8}^2 (X_{n-3} - X_{n-2}) / X_{n-3}X_{n-6}$$

(d) 방정식

<그림 3> RR 간격 모델링 결과

4. 토의

실험 결과, 적합도 평가 함수에서는 신호값의 분산, 구간별 히스토그램, fractal dimension의 의미가 있었다. 유전자의 수와 최대 반복 세대수를 변화시켜 보았을 때, 유전자의 수가 많고 최대 반복 세대수가 증가할 수록 정확한 모델링 방정식을 찾아내었다. 유전자 프로그래밍의 실행 시간은 적합도 평가 함수의 복잡도에 비례하며 시간 복잡도가 높은 correlation dimension을 포함한 경우에는 실행 시간이 현저히 증가하였다. SUN Sparc 20에서 실행하여 적합도 평가 함수로 시계열 데이터의 차이, 평균의 차이, 분산의 차이, 구간별 히스토그램, 값의 범위의 차이를 사용하였을 때, 유전자의 수는 200개, 최대 반복 세대수는 100인 경우 모델링 시간이 50분 정도 소요되었다. 적합도 판정 함수에 fractal dimension을 추가한 경우에는 모델링에 90분정도 소요된다. 시간 복잡도가 높은 correlation dimension을 추가하는 경우에는 3일 이상 소요되었으나, 유전자 프로그래밍의 모델링 성능에는 큰 차이는 없었다. 따라서 적합도 평가 함수로는 통계적 특성과 fractal dimension이 적합함을 알 수 있다.

생체 신호의 경우에는 잡음이 포함되어 있는 시계열 데이터를 사용하였으며, 모델링한 방정식으로부터 시계열 데이터를 생성하여 생체 신호와 비교하면 통계적 특성과 혼돈 특성이 유사한 결과를 보였다.

참고 문헌

- [1] Alister I. Mees, "Tesselations and dynamical systems", Proc. of the Workshop on Nonlinear Modeling and Forecasting, vol. XII, pp. 3-24, 1990.
- [2] W. C. Mead, R. D. Jones, Y. C. Lee, C. W. Barnes, G. W. Flake, L. A. Lee, M. K. O'Rourke, "Prediction of Chaotic Time Series using CNLS-Net-Example: The Mackey-Glass Equation", Proc. of the Workshop on Nonlinear Modeling and Forecasting, vol. XII, pp. 39-71, 1990.
- [3] J Gleik, "Chaos : Making a New Science", New York: Viking Penguin Inc., 1987.
- [4] T.S. Parker, L.O. Chua, "Chaos: A Tutorial for Engineers.", Proc. IEEE, vl. 75, no. 8, pp. 982-1008, Aug, 1987.
- [5] E.N. Lorenz, "Deterministic Nonperiodic Flow", J. Atmospheric Sci., vil. 20, no. 2, Mar., 1963.
- [6] R. Pool, "Is It Chaos, or Is It Just Noise",

Science, vol.1.243, pp. 25-28, Jan., 1989.

- [7] R. Pool, "Is It Healthy to Be Chaotic", Science, vol.1.243, pp-604-607, Feb., 1989.
- [8] P. Grassberger, I. Procaccia, "Measuring the Strangeness of Strange Attractors", Physica D, vol. 9, pp. 198-208, 1983.
- [9] A.L.Goldberger, D.R. Rigney, B.J.West, "Chaos and Fractal in Human Physiology", Scientific American, pp. 35-41, Feb., 1990.
- [10] J.P.Eckmann, S.O. Kamphorst, "Liapunov exponents from time series", Physica Review A, vol. 34, no. 6, pp. 4971-4979, Dec., 1986.
- [12] D. Gulick, "Encounters with Chaos". McGraw-Hill Inc., 1992.
- [13] N. Packard, J. Crutchfield, D. Farmer, R.Shaw, "Geometry from a time series", Phys. Rev. Lett., vol. 45, pp 712-715, 1980.
- [14] D.S. Broomhead, G.P.King, "Extracting qualitative dynamics from experimental data", Physica D, vol. 20, pp. 217-236, 1986.