

유전자 프로그래밍을 이용한 비선형시스템 모델링에 관한 연구

김보연, ¹박광석

서울대학교 협동과정 의용생체공학과

¹서울대학교 의과대학 의공학교실

A Study on the Modeling of Nonlinear System Using Genetic Programming

B. Y. Kim, K. S. Park

Interdisciplinary Program of Med. and Bio. Eng. Major, Seoul Nat'l University

Dept. of Biomedical Engineering, College of Medicine, Seoul Nat'l University

Abstract

Even though there are several deterministic methods for the modeling of linear systems, there is no standard method for the modeling of nonlinear systems. For the modeling of nonlinear systems we have applied the genetic programming method to estimate nonlinear time series. We get the time series from the simple known nonlinear dynamics, and fed those to genetic programming. For the tested nonlinear systems, suggested method estimated the nonlinear dynamics correctly.

1. 서론

비선형 시스템에 대한 모델링 방법은 다양하게 연구되고 제안되고 있다. 그러나 주어진 비선형 신호에서 이를 정의하는 방정식인 해를 찾기 위해서는 거대한 해 공간(solution space)을 탐색(search)해야 하므로 이를 찾아내는 것은 쉽지 않다. 본 연구에서는 주어진 비선형 신호에 대한 정확한 모델링 방법으로 유전자 프로그래밍을 사용하였다. 유전자 프로그래밍 기법은 거대한 탐색공간에서 임의의 해를 추출하여 이를 평가한 후에 적합한 해 공간을 선택하여 다시 탐색을 계속해 나가는 방법이다. 이렇게 함으로써 방대한 탐색공간을 좁혀나가면서 해를 찾아나가는 데 유전자 프로그래밍 방법으로 가능한 해의 형태와 크기도 스스로 생성해 낸다. 이 방법으로 생성한 해(solution)가 얼마나 적합한가를 판정하는 것이 가장 중요하며 유전자 프로그래밍 방법의 핵심중의 하나라 하겠다.

본 연구에서는 유전자 프로그래밍 방법이 비선형 시스템 모델링에 적합한 방법여부를 판별하기 위하

여 비선형 시스템으로 알려진 Verhulst의 인구증가 방정식을 적용하였다. Verhulst 방정식의 a 와 초기값을 선택하여 이로부터 비선형 신호를 생성한 후, 이 비선형 신호를 유전자 프로그래밍의 입력으로 사용하였다. 이 비선형 신호를 입력받아 유전자 프로그래밍 기법을 사용하여 다시 Verhulst 방정식을 찾아내는 방식으로 비선형 신호에 대한 모델링을 수행하였다. 이 때, 스스로 찾아내는 방정식의 형태에 제한을 두지 않으면 탐색공간은 무한대로 커지게 된다.

본 연구에서는 비선형 시스템의 예로 기본적인 Verhulst 방정식으로 부터 신호를 생성하여 사용하였으나 유전자 프로그래밍 방법의 입력이 비선형 신호이므로 비선형성을 보이는 생체 신호 및 다른 비선형 신호도 본 연구에서 사용한 유전자 프로그래밍 방법에 적용할 수 있을 것으로 생각한다.

2. 유전자 프로그래밍

2.1 개요

유전자 프로그래밍 방법은 프로그램화되어 있지 않은 문제, 즉 문제를 풀이하기 위해서 필요한 것은 알고 있으나 그 풀이 방법을 알지 못하는 문제를 컴퓨터가 스스로 풀이할 수 있도록 프로그램을 자동 생성해주는 방법이다. 이 방법에서 풀이할 문제를 표현하는 방식, 평가 함수의 사용, 그리고 우수한 유전자의 선택 등은 유전자 알고리즘과 동일한 방법을 사용하고 있다. 즉, 문제의 풀이 방법을 유전자(chromosome)로 표현하고 평가 함수(fitness function)를 사용하여 주어진 유전자의 적합도를 판정하여, 적합도가 높은 유전자를 다음 세대(generation)로 유전시키는 것이다. 그러나 유전자 알고리즘은 문제에 대한 해답을 고정된 형식의 유전자로 표현하는 반면, 유전자 프로그래밍방법으로 생성한 유전자는 고정된 형식이 없다. 어떤 문제를 풀이하기 위한 프로그램의 길이가 길수도, 짧을 수도 있듯이 유전자의 길이도 다양하고 형식도 다양한 형태이다. 즉, 유전자 프로그램은 유전자 알고리

즘의 확장된 방법이라 할 수 있다.

2.2 유전자 프로그래밍의 구성 요소

유전자 프로그래밍을 위하여는 먼저 다음 다음 5가지의 요소를 결정하여야 한다.

- ▶ 터미널의 집합 : 해를 구성하는 변수 및 계수들의 집합
- ▶ 함수의 집합 : 해를 구성하는 함수들의 집합
- ▶ 적합도 판정함수 : 해가 문제에 적합한 정도를 표시하는 척도
- ▶ 수행 제어를 위한 인자
- ▶ 수행을 끝내거나 해를 결정하는 방법

터미널의 집합은 해를 구성하는 변수 및 계수들의 집합으로 비선형 시스템을 표현하는 결정식의 계수 및 변수를 말한다. 즉, 생성될 결정식을 구성하는 요소이다.

함수의 집합은 결정식내에서 사용되는 연산자로서 수치 연산자, 논리 연산자 및 기타 수학적 함수 등이 이에 해당한다. 여기에 포함된 함수도 유전자로 표현되는 결정식을 구성하는 요소가 된다. 따라서 각 결정식은 터미널의 집합 및 함수의 집합의 원소를 조합하여 구성된다.

유전자 프로그래밍에 의하여 생성된 결정식이 문제에 맞는 해(solution)가 되기 위해서는 주어진 요구(requirement)를 만족해야 한다. 이러한 문제에 만족하는 정도를 적합도(fitness)라 한다. 즉, 적합도는 유전자 프로그래밍에서 생성한 결정식이 주어진 비선형 시스템을 얼마나 잘 표현하는가를 수치로 측정한 값이다. 이 적합도에 따라 유전자 프로그래밍이 생성한 결정식들을 취사선택하게 된다. 이 적합도는 여러 가지 평가 방법에 의하여 결정되며 비선형 시스템에 대한 모델링의 정확성을 결정하는 데에는 적합도를 평가하는 함수의 결정이 가장 중요하다.

수행 제어를 위한 인자로서는 발생시킬 유전자(즉, 결정식들)의 수와 교배를 반복하는 최대 세대수(maximum generations)가 필요하다. 이외에도, 유전 연산자를 수행할 확률, 부모 유전자의 선택 방법등의 인자들이 필요하다.

유전자 프로그래밍의 수행이 시작되면 수행의 종료를 결정할 방법도 미리 정해두어야 한다. 예를 들어, 적합도가 어느 값에 인접한 경우도 될 수 있고 최대 세대수 이상 수행이 진행되었을 때로 정할 수도 있다.

2.3 유전자 프로그래밍의 실행

유전자 프로그래밍을 수행하면 미리 정한 수천개의 유전자(본 연구에서는 비선형시스템을 모델화한 결정식)를 생성하여 다음 세대로 유전시키는데 이때 각각의 유전자는 다윈의 적자 생존 원칙에 의하여 선택되어 다음 세대로 유전된다. 유전자 프로그래밍에서 적자 생존의 원칙은 적합도가 높은 유전자를 선택하여 다음 세대로 복사(reproduction)하거나 두 개의 유전자를 선택하여 임의의 위치에서 교배(crossover)하여 새로운 유전자를 생성하여 다음 세대로 넘겨 주는 것이다. 이 때 적합도가 높은 것을 교배하면 다음 세대에 전달되는 자녀들은 더

욱 적합도가 높을 확률이 높다. 이러한 교배 및 전달을 세대를 지나면서 계속 반복하면 마지막에는 적합도가 높은 유전자만이 남게 되어 문제에 가장 적합한 유전자를 구할 수 있다. 이와 같은 방식은 병렬 탐색(parallel search) 방법으로 빠른 시간에 원하는 해를 찾을 수 있는 이점이 있다. 따라서 유전자 프로그래밍 방법을 이용하면 빠른 시간에 문제에 가장 적합한 해를 구할 수 있다.

유전 연산자는 교배(crossover)와 돌연 변이(mutation)의 두 가지가 있다. 교배는 적합도가 높은 두 개의 유전자를 선택하여 유전자 내의 임의의 위치에서 두 개의 조각으로 나눈 뒤, 서로 상대방의 반대 조각을 교환하여 다시 연결하여 하나의 유전자로 재 배열하는 방식이다. 이렇게하면 두 개의 유전자에서 다음 세대의 두 개의 새로운 유전자가 발생된다. 두 개의 부모는 다음 세대에서는 버려진다. 오직 자녀 유전자만을 전달한다. 이 외에도 유전자의 일부를 다르게 대치하는 돌연 변이 방법이 있는데 유전자 집합이 local minima에 빠지는 오류도 막아주며 새로운 형태의 유전자를 생성하여 변화를 주기 위하여 필요하다.

3. 유전자 프로그래밍 방법을 이용한 비선형 시스템 모델링

3.1 문제의 정의 및 실행 방법

유전자 프로그래밍을 이용하여 비선형시스템을 모델링하기 위하여 다음의 과정으로 수행한다.

- 비선형시스템을 선정한다. 본 연구에서는 인구 증가를 수학적으로 표현한 Verhulst의 방정식을 선택하였다. Verhulst의 방정식은 계수와 초기값에 따라서 비선형적인 특성을 보이는 것으로 알려져 있다.

$$x_{n+1} = a \times x_n \times (1 - x_n)$$

위 방정식의 계수 a 와 초기값 x_0 을 선정하여 시간에 따른 x_n 을 생성한다. 생성된 값들은 비선형적이고 비주기적인 특성을 갖는다. 이렇게 생성된 값들을 유전자 프로그래밍으로 생성한 결정식의 적합도 판정에 이용한다.

- 유전자 프로그래밍으로 임의의 결정식을 생성하여 이 식으로 부터 신호를 생성한다. 생성된 식에 대한 적합도를 판정한다. 적합도 판정은 주어진 신호와 결정식에서 생성된 신호를 다양한 방법으로 분석, 비교하여 판정한다.

- 유전자 프로그래밍의 실행이 완료되면서 주어진 비선형 신호에 대한 결정식이 주어진다.

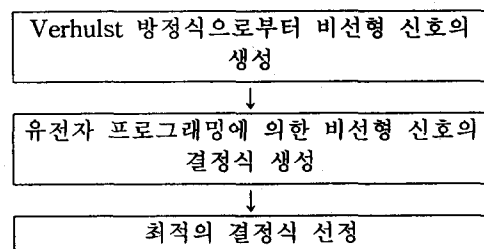


그림 1. 모델링 과정 블록 다이어그램

3.2 문제의 표현

Verhulst의 방정식을 모델링하는 결정식은 다음과 같은 형태로 정한다.

$$X_{n+1} = f(a,b,c,\dots,k,X_n)$$

즉, 비선형 신호를 모델화한 결정식을 유전자로 사용한다. 따라서, 유전자는 다음과 같이 가능한 모든 결정식이 될 수 있다.

$$x_{n+1} = 3.23x_n^2 + x_n$$

$$x_{n+1} = x_n + 1.25^3$$

3.3 유전자 프로그램의 구성 요소

위와 같은 결정식을 유전자 프로그래밍을 통하여 생성하기 위하여 다음과 같이 구성 요소를 정한다.

(1) 터미널의 집합 :

터미널은 1개 이상의 상수(식에서 a, b, \dots, k)와 1개의 변수 x_n 으로 정한다. 상수는 0.0부터 4.0사이의 값이며 소수 둘째자리까지로 제한한다. 상수의 개수에는 제한이 없다.

따라서 유전자 프로그래밍에 의해 생성된 결정식은 0.0부터 4.0사이의 계수 및 상수와 1개의 변수로 구성된다. 변수의 지수 형태에도 제한이 없다. 따라서 찾고자 하는 결정식의 탐색 공간은 무한대이다.

(2) 함수의 집합 :

기본적인 사칙연산(+, -, *, /)만을 허용한다. 지수의 표현은 사칙연산을 이용하여 표시한다.

(3) 적합도 평가 함수 :

적합도 평가 함수는 다음과 같은 항으로 이루어진다.

- ① 두 식에서 생성된 시간별 값의 차이
- ② 평균의 차이
- ③ 분산의 차이
- ④ 값의 구간별 히스토그램의 차이
- ⑤ 값의 범위의 차이

먼저 a 와 x_0 의 값을 선정하여 Verhulst 방정식의 시간에 따른 x_n 을 구한다. 이를 비선형 신호로 사용한다. 시간에 따른 각 x_n 의 값을 기억하고 전체 신호값의 평균과 분산을 구한다. 또한 x_n 은 0.0과 1.0 사이의 값이므로 0.1씩 구간을 정해 각 구간별 값의 히스토그램과 최대, 최소값을 구한다.

유전자 프로그램에 의하여 생성된 결정식은 이미 상수와 계수가 정해져 있으므로 주어진 x_n 의 값을 이용하여 신호를 생성한다. 이 신호도 원신호와 마찬가지로 평균과 분산을 계산하고 각 구간별 히스토그램과 최대, 최소값을 구한다.

두 신호의 각 시간대별 값, 평균, 분산, 구간별 히스토그램, 최대, 최소값의 차이를 모두 합한다. 이 값은 주어진 조건에서의 두 신호가 불일치하는 정도를 나타낸다.

위에서 계산한 값은 주어진 조건에서의 불일

치량이므로 임의의 조건인 경우의 불일치량을 구해 본다. 즉, 임의의 초기값에 대하여 원래의 식과 생성된 결정식으로 다시 신호를 생성하여 위에서 구한 불일치 정도를 구한다. 5개의 임의의 초기값에 대하여 신호를 생성하고 불일치 정도를 구하여 위에서 구한 불일치 정도에 합한다.

이 모든 합이 두 식의 총 불일치 정도가 되며 이 값의 역수가 적합도가 된다.

(4) 수행 제어를 위한 인자 :

유전자의 수는 20 ~ 200개, 최대 세대수는 50으로 하며 교배확률은 0.3, 돌연변이 확률은 0.7로 한다. 돌연변이 확률을 높이면 local minima에서 나갈 수 있는 가능성이 높아진다.

(5) 수행을 끝내거나 해를 결정하는 방법:

오류가 0이거나 최대 세대수 이상 유전하면 수행을 종료하고 가장 적합도가 큰 것을 주어진 비선형시스템을 모델링한 결정식으로 정한다.

4. 실험 및 결과

다음과 같은 Verhulst 방정식

$x_{n+1} = a \times x_n \times (1 - x_n)$ 의 a 와 x_0 를 정하여 이로부터 비선형 신호를 생성하여 이 신호를 모델링한 3가지 경우를 보인다.

(1) 경우 1: 2주기 신호

Verhulst 방정식에 $a=3.00$ $x_0=0.9$ 로 신호를 생성하면 그림 2와 같은 2주기 신호가 생성된다.

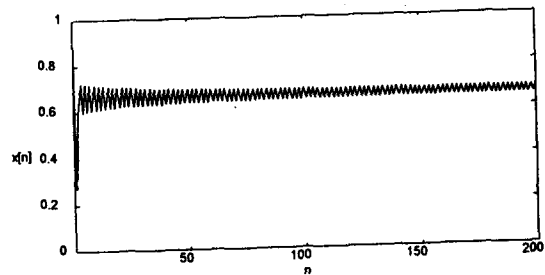


그림 2. 2주기 신호

실험에 사용한 인자는 다음과 같다.

- 한 세대에서의 유전자의 수 : 1000개
- ▷ 반복 세대 수 : 50 세대
- ▷ 입력에 사용한 신호의 포인트 수 : 100포인트
- ▷ crossover rate : 0.3
- ▷ mutation rate : 0.7

평가 함수는 ① 두 신호의 시간별 값의 차이

- ② 평균의 차이
- ③ 분산의 차이
- ④ 값의 구간별 히스토그램의 차이
- ⑤ 값의 범위의 차이

를 사용하였다.

유전자 프로그래밍으로 위의 신호를 모델링한 결과

$$x_{n+1} = 3.00 \times x_n \times (1 - x_n)$$

유전자 프로그래밍을 이용한 비선형시스템모델링에 관한 연구

으로 신호를 생성한 원식과 일치하는 방정식을 찾아냈다.

이때, 탐색공간에서의 global minimum을 찾기 위하여 mutation rate를 높였고, 최적의 유전자는 10세대를 반복하면서 적합도가 가장 높았던 유전자로 선택하였다.

(2) 경우 2: 4주기 신호

$a=3.62$, $x_0=0.7$ 인 경우에는 그림 3과 같은 4주기 신호가 발생된다.

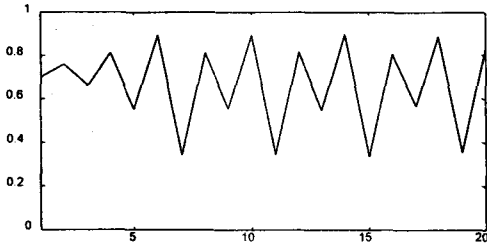


그림 3. 4주기의 신호

실험에 사용한 인자는 다음과 같다.

- ▷ 한 세대에서의 유전자의 수 : 1000개
- ▷ 반복 세대 수 : 10 세대
- ▷ 입력에 사용한 신호의 포인트 수 : 20포인트
- ▷ crossover rate : 0.3
- ▷ mutation rate : 0.7

평가 함수는 ① 두 신호의 시간별 값의 차이

- ② 평균의 차이
- ③ 분산의 차이

를 사용하였다.

유전자 프로그래밍으로 위의 신호를 모델링한 결과

$$x_{n+1} = 3.62 \times x_n \times (1 - x_n)$$

으로 신호를 생성한 원식과 일치하는 방정식을 찾아냈다.

(3) 경우 3: 혼돈 신호

$a=3.75$, $x_0=0.5$ 인 경우에는 그림 4와 같은 혼돈 신호가 발생된다.

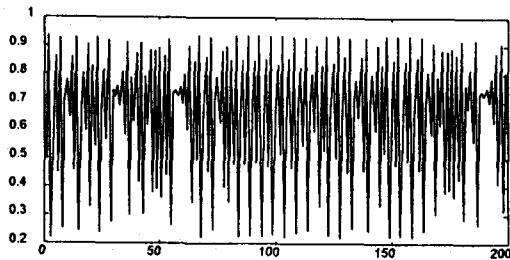


그림 4. 혼돈 신호

평가 함수 및 실행 제어 인자는 실험 1과 동일하게 사용하였다.

유전자 프로그래밍으로 위의 신호를 모델링한 결과

$$x_{n+1} = 3.75 \times (x_n - x_n^2)$$

으로 신호를 생성한 원식과 일치하는 방정식을 찾아냈다.

5. 토의

비선형 시스템에 대한 모델링 방법으로써 유전자 프로그래밍 방법을 이용하였다. 실험 결과, 유전자 프로그래밍 방법은 비선형 시스템 모델링에 적합한 방법이며, 적합한 평가 함수도 제안하였다. 평가 함수에서는 신호값의 분산, 구간별 히스토그램, 값의 범위가 의미가 있었다. 실험을 제어하는 인자인 crossover rate와 mutation rate를 비교해보면, mutation rate가 클수록 정확히 모델링하였고, 반복 세대수가 크고, 세대당 유전자의 수가 클수록 정확한 결정식을 찾았다.

본 연구에서 시도한 방법은 비선형 신호만을 가지고 이를 모델화한 방정식을 찾아내는 방법이었다. 흔히 자연계에서 보이는 신호들, 생체 신호들은 비선형적인 특성을 보이며, 이로 부터 얻을 수 있는 정보도 신호외에는 거의 없으므로 본 연구 방법을 비선형적인 특성을 보이는 다양한 신호들에 응용할 수 있으리라 생각한다. 이를 위해서는 본 연구에서 제한했던 유전자 표현 방식을 확장하고, 새로운 시스템에 적합한 평가함수를 수정하고 혼돈특성을 측정하는 인자를 추가하는 연구를 진행중이다.

6. 참고문헌

- [1] James Gleick, "Chaos", 1987, Penguin Books
- [2] Nicholas B. Tufillaro, "An Experimental Approach to Nonlinear Dynamics and Chaos", 1992, Addison-Wesley Publishing Company
- [3] S. Neil Rasband, 1990, "Chaotic Dynamics of Nonlinear Systems", A Wiley-Interscience Publication
- [4] Lawrence Davis, 1991, Handbook of Genetic Algorithms, Van Nostrand Reinhold
- [5] John R. Koza, 1994, Genetic Programming II: Automatic Discovery of Reusable Programs, The MIT Press