

# 시계열 예측을 위한 DNA 코딩 방법

## DNA Coding Method for Time Series Prediction

°이 기 열, 선 상 준, 이 동 옥, 심 귀 보

중앙대학교 전자전기공학과 (Tel: 02-820-5319; Fax: 02-817-0553; E-mail: kbsim@cau.ac.kr)

**Abstract :** In this paper, we propose a method of constructing equation using bio-inspired emergent and evolutionary concepts. This method is algorithm that is based on the characteristics of the biological DNA and growth of plants. Here is, we propose a constructing method to make a DNA coding method for production rule of L-system. L-system is based on so-called the parallel rewriting mechanism. The DNA coding method has no limitation in expressing the production rule of L-system. Evolutionary algorithms motivated by Darwinian natural selection are population based searching methods and the high performance of which is highly dependent on the representation of solution space. In order to verify the effectiveness of our scheme, we apply it to one step ahead prediction of Mackey-Glass time series.

**Keywords :** DNA Coding Method, Time Series Prediction, L-system, Biological DNA, Genetic Algorithms(GA)

### 1. 서론

자연계에는 많은 생물들과 이 생물들이 서로 상호 작용을 하며 살아가는 생태계가 있다. 이런 생물들의 삶을 모방한 많은 인공생명의 모델들이 있다. 생물은 자신이 가진 유전자를 바탕으로 발생하고, 이렇게 발생된 개체는 생존해 나가며, 학습을 한다. 또한 자신의 유전 정보를 후손에게 전달함으로써 종족을 보존한다. 이러한 생물의 자기 조직화 현상을 모방한 인공생명의 모델은 진화모델, 발생모델, 발생/발달 모델 3가지가 있다. 진화모델로는 진화 알고리즘, 학습모델로는 신경망, 강화학습, 인공면역계 등이 있으며, 발생/발달모델로는 셀룰러 오토마타(CA) 린든마이어 시스템(L-system)[1~3] 등이 있다.

본 논문에서는 원하는 목적의 신경망을 얻기 위하여 DNA 코딩 [4]을 이용하여, 규칙을 생성시킨 후 그 규칙을 통해 수식을 구성한다. 또한 진화알고리즘을 이용하여 점점 더 우수한 개체를 선택함으로써 최종적으로 원하는 목적의 수식을 구현한다.

DNA 코딩은 생물학적인 DNA 구조를 모방한 것으로서 DNA가 생물의 유전정보를 통해 자신을 발생시키고, 또한 다음 세대에 유전 정보를 전달하는 과정을 모방한 방법이다. 유전자에서는 같은 코돈(유전자의 최소 단위, 3개의 염기 배열로 구성)일지라도 코돈의 위치나 좌우의 상관 관계에 의하여 다른 형질로 발현되는 특징을 갖는다. 이러한 특성을 모방한 DNA코딩은 동적인 구조를 통한 중복 해석과 여분이 있다는 장점을 이용하여 하나의 해석 단위를 신경망의 노드와 그의 부수적인 요소(Weight, 입력력 범위)를 결정하게 한다. 그리고 중복해석을 통해 나온 L-system의 생성규칙을 이용하여 구성)을 이용하여 수식을 결정한다.

이렇게 자동 생성된 수식을 Mackey-Glass 시계열 예측 문제에 적용시켜, 우수한 개체를 선택한 후 이들 개체의 유전자를 GA 연산자(돌연변이와 교배)를 통하여 다양화된 유전자를 다음 세대로 전달하고, 더 좋은 개체를 얻어 시계열 예측문제를 풀 수 있는 신경망을 자동적으로 생성하도록 하는 것이다.

### 2. DNA 코딩 방법

#### 2.1 생물학적 DNA

모든 생물체는 각자 고유의 DNA를 가지고 있다. DNA는 개체의 특성을 발현시키는 유전코드로서, A(아데닌) T(티민, RNA에서는 U:우라실) G(구아닌) C(시토신)의 4개의 염기배열로 이루어져 있다. 또한 염기 3개의 배열이 한 의미단위를 이루어 해석된다. 이 의미단위를 생물학적인 용어로 코돈(codon)이라 한다. 코돈의 가짓수는  $4 \times 4 \times 4 = 64$ 개이며 이것이 코드화 하는 아미노산은 20가지이다.[5].

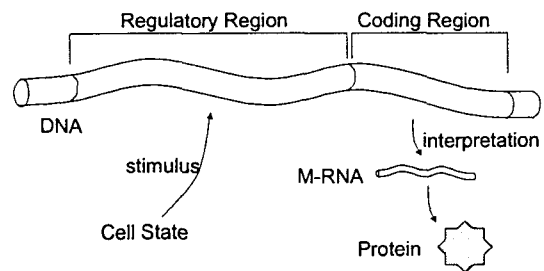


그림 1. 유전자 구조의 일반적인 구조.

생물학적 유전자의 기본적인 구조는 그림 1과 같이 단백질을 직접 코드화 하는 부위(coding region)와 그 코드 부위의 발현을 조절하는 조절부위(regulatory region)로 구성되어 있다. 코드화 부위는 조절부위의 명령에 의해 세포내의 조건이 조절부위를 자극할 때 단백질로 번역된다. 이것은 발생모델의 규칙의 표현방식과 유사하다. 즉, 조절부위는 규칙의 전건부 또는 세포의 현재 주변 상태, 코드화 부위는 후건부 또는 세포의 다음 상태에 대응된다.