

DNA Coding 및 L-system에 기반한 진화신경회로망

Evolutionary Neural Networks based on DNA coding and L-system

이기열, 전호병, 이동욱, 심귀보

중앙대학교 전자전기공학부

Tel : 02) 820-5319, kbsim@cau.ac.kr

Ki-Youl Lee, Ho-Byung Chun, Dong-Wook Lee and Kwee-Bo Sim
School of Electrical and Electronics Engineering,
Chung-Ang University

In this paper, we propose a method of constructing neural networks 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

I. 서론

자연계 생명체는 자기 종족의 특성을 보존하고, 또한 환경에 적응하면서 종족을 보존, 번성하고 있다. 이런 자연계 생명체의 특징을 공학적으로 이용하려는 노력은 단순한 행동의 모방에서 시작을 하여 성장, 유전, 번역, 진화에 이르기까지 다양한 분야로 나뉘어져 연구 되고 있다. 그 대표적인 분야로 진화를 모방한 유전자 알고리즘, 유전자를 이용한 DNA Coding, 식물의 성장을 모델링한 L-system, 번역계를 모방한 인공 번역계, 그리고 뇌의 뉴런을 이용한 뉴럴 네트워크 등 다양한 연구 분야가 있다.

최근 신경회로망의 구조나 파라미터를 진화시키기 위하여 진화적 접근방식이 많이 도입되고 있다. 또한 결과 신경회로망은 일반적으로 백프로퍼게이션을 통하여 학습된다. 따라서 이 경우 계산비용이 너무 높아 유전자 알고리즘을 적용하기에 작은 규모의 구조

를 최적화 하는 경우를 제외하고는 거의 불가능하다.

이러한 문제를 보완한 DNA 코딩과 L-system의 방법을 이용한 신경회로망의 구조를 제한한 바 있다.

본 논문에서는 유전형과 표현형의 번역을 L-시스템 같은 발생모델을 이용한 성장과정을 이용하였다. 이 때 염색체는 신경회로망의 구조를 코드화 하는 것이 아니라 생성 규칙을 코드화 하기 때문에 신경회로망의 크기가 커지더라도 염색체의 길이는 길어지지 않는다. 이러한 발생 모델에 기반한 신경회로망의 구조는 모듈성과 자기 유사성을 갖는다.

본 논문에서는 L-시스템 규칙의 진화시키기 위한 DNA코딩방법을 제안한다. DNA코딩은 중복과 여분을 효과적으로 사용하여 규칙의 표현에 적합하다. 본 논문에서는 L-시스템과 일반적인 표현방법을 개발하였다. 이 두 가지 발생모델과 코딩방법을 사용하여 L-시스템 신경회로망과 개발하였다.

L-시스템 기반 신경회로망은 모듈성과 실수 연결강도를 갖는다. 따라서 이 신경회로망은 예측, 제어, 인식, 분류 등과 같은 실수치 문제에 효과적이다. L-시스템 신경회로망의 유용성을 검증하기 위하여 Mackey-Glass 데이터의 시계열 예측문제에 적용하였다. 또한 일반적인 백프로퍼게이션을 이용한 신경망과 비교하여 유효성을 입증한다.

II. DNA 코딩

DNA의 동작을 모방한 코딩방법에 대한 연구는 Yosikawa[4] 등에 의하여 기본적인 방법이 연구되었다. 염색체는 기본적으로 4가지 염기의 배열로 이루어져 있고 아미노산을 번역하는 것과 마찬가지로 코돈 단위로 번역한다. 또한 다음 그림 2와 같이 유전자의 번역 시작점이 일정하지 않기 때문에 중복 번역을 허용하며 때에 따라서는 번역되지 않는 부분도 존재한다. 하나의 아미노산을 생성하는 코돈이 여러 개이기 때문에 염색체의 중복을 효율적으로 이용할 수 있다. 교차점도 임의로 주어지기 때문에 염색체의 길이도 가변적이다. Yosikawa의 방법은 번역 테이블을 만들므로서 규칙의 표현에 적합하게 구성되어 있다. 이상의 특징을 정의하면 아래와 같은 4가지로 정리할 수 있다.

- ◆ DNA 코딩방법의 특징
 - (1) 지식의 표현이 쉽다.
 - (2) 코딩에 여분과 중복이 있다.
 - (3) 염색체의 길이가 가변적이다.
 - (4) 교차점에 제약이 없다.

CGATG CCG CGT CAT GAA TGC CGG GGT TC CAT ACC TCG GGA C
 Arg Arg His Glu Cys Arg Gly
 Pro Gly Phe His Thr Ser Gly

그림 1. DNA 염색체의 번역.

그림 2에서 보듯이 위와 같은 DNA코드가 있다면, 그 안에 있는 시작코돈(ATG)부터 해석을 시작하여 각각의 아미노산을 해석하여 그에 따른 규칙을 생성한다. 이와 같은 해석 방법을 통하여 짧은 DNA코드에서도 많은 정보를 얻을 수 있다.

III. L-System

L-system[1~3]은 병렬적인 문자열의 재조합 속

성을 갖는 일종의 문법으로서 1968년 Aristid Lindenmayer에 의해 제안되었다. 이것은 다세포 생물의 성장 과정의 모델링이 가능하기 때문에, 이후 컴퓨터 그래픽 등에서 식물을 모델링하는데 많이 이용되고있다.

IV. 신경망의 DNA 코딩방법

4.1 신경망 구성을 위한 DNA 코딩방법

본 절에서는 신경망을 진화시키기 위한 DNA코딩방법을 설명한다. 신경망을 구성하기 위해 DNA 코드에서 규칙(L-system의 규칙)을 생성한 후, 그 규칙을 바탕으로 신경망을 구성한다.

우선 임의의 DNA 코드를 발생시킨다. 이 코드를 위 표에 의하여 해석을 하여 여러 개의 규칙을 만든다. 시작코돈(ATG)이 나오면 해석을 시작한다. 처음에 나오는 첫 코돈은 신경망의 문자(아미노산을 그대로 이용)로 해석을 한다. 두 번째 코돈은 신경망의 연결 범위(connecting range)를 결정한다. 예를 들어 숫자가 3이라면 최종적으로 만들어진 문자열에서, 그 노드로 3번째 노드까지 모두 연결이 된다. 단, 쉼표(,)가 나오면 그 옆의 노드는 연결하지 않는다. 세 번째 코돈은 신경망의 노드에 포함되는 Bias로 해석을 한다. 값의 계산은 DNA A=0, G=1, T=2, C=3으로

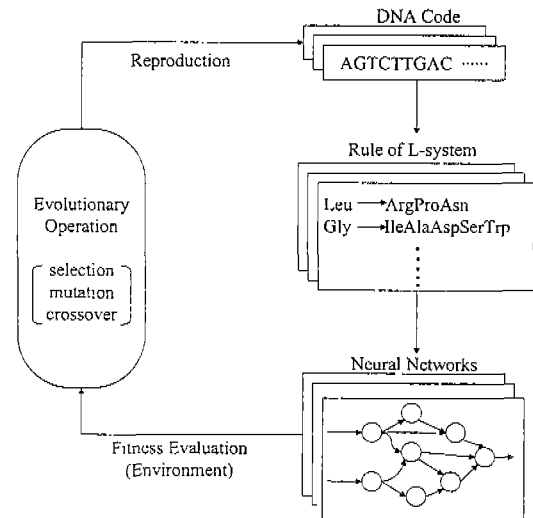


그림 2 전체 다이어그램

하여 식 (1)과 같이 계산을 한다. 이 후에 나타나는 코돈은 두 번째 코돈에서 정해진 숫자만큼 Weight 값

1) Connecting Range (x,y) : 문자열(노드의 배열)에서, 현재 노드에서 연결할 수 있는 범위란 x 번째 노드부터 y 번째 노드까지 정함. 단 쉼표(,)가 나올 경우 쉼표 뒤에 있는 노드는 연결하지 않으나 연결 범위에는 들어감.

으로 해석을 한다. 계산은 Bias와 같은 방법으로 한다. 여기까지 해석을 하면 하나의 신경망노드를 구성한 것이다. 같은 방법으로 정지 코돈이 나올 때까지 해석을 계속한다.

$$Bias = \frac{(DNA \times 4^2 + DNA \times 4^1 + DNA \times 4^0) - 32}{10} \quad (1)$$

(-3.2 ≤ W ≤ 3.2 0.1간격)

그리고 종료코돈이 나올 때까지 번역을 하여 아미노산의 배열을 만든다. 각각의 규칙에 대해 처음 나온 코돈의 아미노산을 규칙의 전건부로 나머지를 후건부로 구성을 한다. 20개의 아미노산이 존재하고, 중복을 허용하지 않으므로 최대 20개의 틀을 형성할 수 있다.

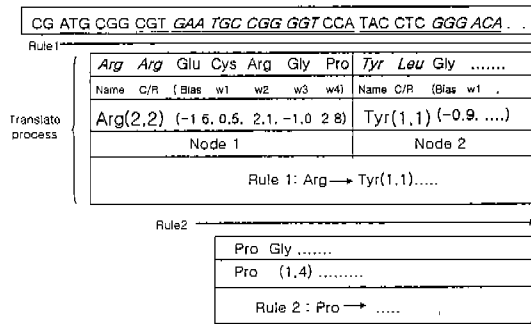


그림 3 DNA의 해석

이렇게 해석된 규칙은 문맥자유 L-system의 규칙과 같은 특성을 지닌다. 그 규칙을 가지고, 정해진 초기 문자열에 따라 규칙을 적용해 그 다음 단계의 문자열을 만들고, 맞는 규칙이 없을 경우에는 그 문자를 그대로 유지한다. 정해진 수 만큼 반복하여 문자열을 치환한 후 생긴 최종 문자열(아미노산의 배열)을 구성한다.

문자열(아미노산의 배열)에서 하나의 아미노산은 표 2. 와 같은 노드와 연결 범위를 갖는다.

이렇게 생성된 신경망은 반드시 한 개 이상의 입력노드와 한 개 이상의 출력노드를 갖는 무정형의 신경망이 된다. 문제에 따라 적당한 수의 입력노드와 출력노드를 갖는 신경망을 선택하여, 주어진 입력을 넣고, 그 출력을 검사하여 원하는 결과와 비교한 후, 유전자 알고리즘에 의해 신경망을 진화시킨다. 신경망의 적합도는 식 (2), (3)에 의해 구한다.

$$Fitness = e^{-\lambda \cdot E} \quad (2)$$

$$E = \sum_{i=1}^n (R_i - O_i)^2 \quad (3)$$

(R: desired output, O: output, E: error, λ: coefficient)

여기서 λ는 비례상수이며, R은 원하는 출력, O는 신경망의 출력이다.

이 집단을 진화시키기 위하여 교배와 돌연변이, 진화전략(Evolution Strategy)의 (μ+λ)선택 방법을 사용하였다. 생성된 신경망의 DNA를 돌연변이와 교배를 통하여 3배수의 자손을 생성한 뒤, 원래 부모 세대의 신경망과 같이 평가하여 순위선택으로 우수한 개체를 뽑아 다음세대의 부모 개체로 삼는다. 이 과정을 반복하여 점점 좋은 개체를 얻는다.

이 방법의 유효성을 검증하기 위해 Mackey-Glass 시계열 예측문제를 해결할 수 있는 수식을 구성하였다.

다음의 예를 통해 위 과정을 설명한다. 다음과 같은 Axiom과 4개의 규칙이 생성되었고, 3번의 반복 생장을 하여 수식을 구성하는 예이다.

Example :

Axiom : Leu

Rule : R1 : Leu → SerProArgGly

R2 : Arg → ArgPhe

R3 : Phe → IleAsn

Step : Step 1 : SerProArgGly

Step 2 : SerProArgPheGly

Step 3 : SerProArgPheIleAsnGly

표 2. DNA 코드표

Amino Acid	# of Amino Acid	Node's Name	Connecting Range
Leu	6	N	1,1
Arg	6	N	2,2
Ser	6	N	3,3
Thr	4	N	1,2
Ala	4	N	1,3
Gly	4	N	1,4
Val	4	N	2,3
Pro	4	N	2,4
Stop	3		
Ile	3	N	3,4
Tyr	2	N	4,4
Gln	2	N	1,1
Phe	2	N	2,2
Asp	2		3,3
Cys	2		1,2
Asn	2		1,3
Glu	2		1,4
His	2		2,3
Lys	2		2,4
Trp	1	N	3,4
Met	1	N	4,4

V. Mackey-Glass 시계열 예측문제

Mackey-Glass 함수는 카오스시스템의 대표적인 예로 식 (5)와 같이 표현된다.

$$\frac{dx(t)}{dt} = \frac{bx(t-\tau)}{1+x(t-\tau)^c} - ax(t) \quad (5)$$

$x(t)$ 는 약 3.5 프랙탈 차원의 스트레인지 어트랙트(strange attractor)를 가지는 카오스 함수이다. $a = 0.2$, $b = 0.1$, $\tau = 1.7$ 로 한 경우의 형상을 시뮬레이션에 사용을 하였다.

Mackey-Glass 시계열 예측 문제의 경우는 DNA 길이 500으로 무작위로 발생된 초기개체군을 시작으로 진화를 시작하였다. 교배율은 0.9, 돌연변이율은 0.3으로 설정하였다. 교배방법은 일점(one point)교배 방식을 사용하였고, 2개 부모개체의 DNA에서 같은 위치에서 교차가 이루어지게 하여 DNA의 길이는 유지하도록 하였다. 그리고 개체의 선택방법은 Ranking Selection과 ES의 $(\lambda + \mu)$ 선택법을 혼합하여 선택하였다.

신경망의 입력은 과거 데이터 값으로 $x(t)$ 의 값을 구하는데 $x(t-1) \sim x(t-19)$ 의 값 중에서 구성된 신경망의 입력노드 수만큼 사용하였으며, 신경망 노드에서의 출력함수는 식 (6)을 사용하였다.

$$f(i) = \left(\frac{2}{1 + e^{-\eta_i}} - 1 \right) \times 2 \quad (6)$$

(i : input η : coefficient)

입력노드는 5개에서 19까지 가질 수 있으며, 출력노드는 하나만 갖는 신경망을 평가 대상으로 하여 200개의 데이터를 넣고, 출력 값을 계산하여 오차가 적은 신경망을 선택하여 다음 세대의 부모 개체로 삼는다.

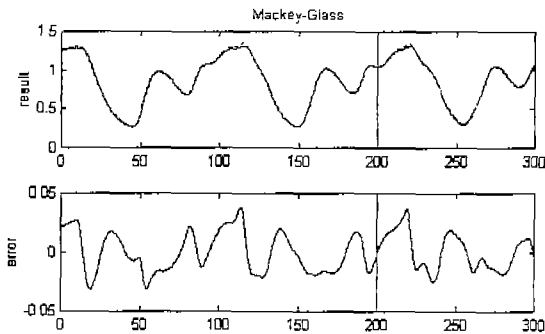


그림 4 Mackey-Glass 시계열 예측 결과 - 진화신경망
(... ideal - predicted)

다음은 입력노드 7개, Hidden 노드 10개, 출력 1개인 신경망을 구성하여, 백프로퍼게이션 알고리즘을

이용하여 학습을 한 결과이다.

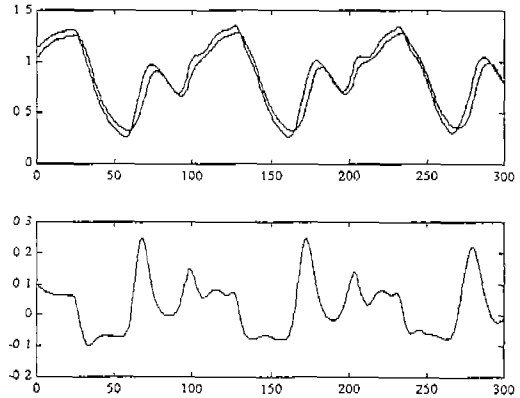


그림 5 Mackey-Glass 예측 결과 - 일반 신경망
(... ideal, — predicted)

VI. 결론

이 논문에서는 발생/발달 모델의 하나인 L-system과 DNA 코딩 방법을 이용하여 신경망을 진화시키는 방법을 제안하였다. DNA 코딩을 이용함으로써 L-system의 생성규칙을 짧은 DNA로 표현하였고, 생성된 규칙을 바탕으로 문자열을 구성한 후 이를 신경망으로 구성하였다. 이렇게 구성된 신경망을 Mackey-Glass 시계열 예측문제에 적용하여 보았다.

이렇게 규칙을 통해 신경망을 구성함으로써, 작은 DNA의 변화로 신경망 전체에 큰 변화를 유도 할 수도 있다. 또한 규칙의 수가 많아짐에 따라 약간의 수렴이 늦어지는 볼 수 있다. 그러나 DNA 코딩에서 L-system 규칙으로 변환에 일반화된 방법을 제공할 수 있다.

감사의 글

본 연구는 과학기술부의 뇌과학 프로젝트(Brantech21)의 지원으로 이루어진 결과임 (과제번호 : 98-J54-01-01-A-07)

참고문헌

- [1] P. Prusinkiewicz, M. Hammel, J. Wolters, R. Mech, "Visual Models of Plant Development," *Hand Book of Formal Languages*, Springer-Verlag, 1996
- [2] Aristid Lindenmayer, Prezemyslaw Prusinkiewicz, "Developmental Models of Multicellular Organisms : A Computer Graphics Perspective," *Artificial Life VI*, pp. 221-249, 1987
- [3] Aristid Lindenmayer, "Mathematical Models for Cellular Interaction in Development, Part I, II," *Journal of Theoretical Biology*, vol.18, pp. 280-315, 1968.
- [4] T.Yosikawa, T. Furuhashi, "A Combination of DNA Coding Method with Pseudo-Bacterial GA for Acquisition of Fuzzy Control Rules," *Proc. of 1st Online Workshop on Soft Computing*, Aug. pp. 19-30, 1996.
- [5] R.A. Wallace, G.P. Sanders, R.J. Ferl, *BIOLOGY : The Science of Life 3rd eds.*, HarperCollins Publishers Inc., 1991.