

## DNA Coding Method에 기반한 신경회로망 진화 기법

\* 이 원희\*, 강훈\*\*

\*중앙대학교 제어계측학과

Tel:+82-02-816-8234; Fax: +82-02-816-1865

\*\* 중앙대학교 전자전기공학부

Tel :+82-02-820-5320; Fax: +82-02-816-1865

### Neural Network Evolution based on DNA Coding Method

\* Dept. of Control and Instrumentation Eng. Chung-Ang Univ

E-mail: wony@sirius.cie.cau.ac.kr

\*\* School of Electronics and Electrical Eng. Chung-Ang Univ.

E-mail :hkang@cau.ac.kr

#### Abstract

In this paper, we propose a new neural network based on the DNA coding method. The initial population of the structure information and the weights for the neural network is generated, and then the descendants are chose with the Elitist selection by the genetic algorithm. The evolutionary technique and the suitable fitness measure are used to find a neural network with the fractal number of layers, which represents a good approximation to the given function.

#### I. 서론

최근 여러 분야에서 신경회로망의 이용은 급속도로 증가하고 있다. 하지만 사용되는 신경회로망은 특정 문제에 잘 동작하도록 특성화된 상태로 설계가 되며 적용되는 문제가 달라짐에 따라 다시 복잡한 설계과정을 거쳐 수정되어야 한다. 이러한 신경회로망의 설계상 난점을 보완하기 위해, 많은 진화적 기법이 도입된 신경회로망 설계모델이 제시되고 있으며 만족할만한 성과를 보이고 있다. 하지만 이러한 모델들도 적용하는 문제가 달라지며 진화 기법을 다시 설계해야 하는 문제점을 안고 있다.

자연계에 존재하는 생물은 DNA라고 하는 자신의 발 생 정보를 가지고 있는 유전물질을 가지고 있다. 이러한 DNA를 통해 자신을 발생시키고 다음세대에 이러한 발생정보를 유전시킨다. 이러한 과정 속에서 생물은 진화한다. 이런 진화 과정을 공학적으로 도입한 것이 진화 알고리즘이다. 하지만 기존의 진화 알고리즘은 개체의 유전정보를 정화된 코드로 설계해야 하는 문제점이 있다. 유전정보의 설계가 잘못되면 진화하지 않는 문제점을 나타낸다. 이러한 문제점을 줄이기 위

해 Yoshikawa가 제시한 모델이 DNA Coding Method<sup>[1][2]</sup>다.

본 논문에서는 DNA Coding Method를 신경회로망 자동설계 및 진화에 응용해 보았다. 그림 1과 같이 하나의 chromosome 속에 포함된 Gene들은 각각 그것에 상응하는 Node들을 발생시키며 발생된 node들의 조합으로 하나의 신경회로망이 구성된다. 0

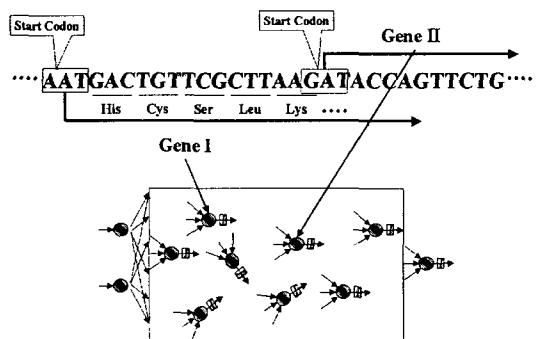


그림 1 신경회로망 발생  
Fig 1 Neural Network production

#### II. DNA Coding method 소개

DNA Coding Method는 1995년 Yoshikawa에 의해 제시된 변형된 형태의 genetic algorithm이다. 일반적인 GA에서 사용되는 이진표현법이 아닌 A(adanine), G(guanine), C(cytosine), T(thymine)로 구성된 네 개의 Symbol을 사용하여 Artificial DNA를 구성한다. 위와 같이 구성된 chromosome은 그것을 포함한 개체의

유전 정보를 포함하고 있으며 이 유전정보는 crossover와 mutation이라는 일반적인 SGA<sup>[3]</sup>연산자에 의해 변형되고 재생과정을 통해 다음 세대에 전달된다. 전달된 유전 정보는 translation mechanism에 의해 다시 새로운 개체를 생성하며, 생성된 개체는 그 자신의 fitness 환경에 의해 평가되고 그것의 적합도에 따라 다음세대에 전파된다. 위의 유전단계는 생물학적 유전과 동일하고 일반적인 SGA와도 동일하다.

Artificial DNA를 구성할 때 고려할 점은 적절한 종류의 start codon 사용과 translation mechanism이며, chromosome내에서 nucleotide가 어디에 위치해 있는가 보다는 Gene상에서 codon이 어떻게 배열되어 있는가로써 정보를 표현하기 때문에 chromosome의 길이를 가변적으로 할 수 있으며, 이러한 가변적인 특성 때문에 crossover point에도 구애받지 않는다.

생물학적 DNA에서 세 개의 nucleotide가 하나의 amino acid로 번역되는 codon이란 단위를 이룬다. 따라서 세 개의 nucleotide로 이루어질 수 있는 codon의 종류는 64가지이며, 인간의 경우 64가지의 codon이 20 가지의 amino acid를 생성한다. 생성된 amino acid들은 배열순서에 따라 여러 가지의 단백질을 구성하며 이러한 단백질들의 조합으로 하나의 세포를 구성한다. artificial DNA의 translation mechanism도 위와 동일한 과정으로 수행된다.

하나의 amino acid를 표현하는데 하나의 nucleotide를 사용하지 않고 codon이라는 단위를 사용함으로써 한 개 이상의 nucleotide를 조합이 동일한 amino acid로 표현함으로써 중요한 유전정보의 과괴를 감소시키는 결실한 성질을 나타낸다. 물론 이러한 성질은 필요 없는 유전정보에도 동일하게 적용되어 수렴속도의 저하를 초래한다.

artificial DNA속에 포함된 Gene은 고정된 위치에 존재하지 않고 그 유전정보의 start codon 이후부터 stop codon 사이에 존재한다. 하나의 Gene 속에 다른 start codon을 포함함으로써 동일한 길이 DNA속에서 가변 개수의 Gene을 생성, 소멸시킬 수 있다.

### III. Neural Network Model 제시

DNA Coding Method를 사용하여 발생시키는 신경회로망은 그 형태가 완전한 무정형의 신경회로망을 목표로 한다. 특별한 layer도 존재하지 않으며 node들의 상호 연결 또한 자신의 위치와 주변상황에 따라 연결이 결정되는 신경회로망을 발생시켜야 할 것이다. 즉 신경회로망내의 어떤 node들간의 연결이 이루어질 수 있으므로 가급적 신경회로망 발생의 제약을 제거하는 모델이 필요하다.

본 논문에서는 위에서 제시한 모델을 다 수용하지는 못하지만 최대한 근접한 모델을 제시했으며 Evolutionary Neural Network(이후 ENNet)이라 부르기로 한다.

ENNet에서는 그림2와 같은 Neural Network Space라는 node가 발생할 수 있는 공간을 정의했다. Neural Network Space는 Layer축과 Node축으로 구성되며 각 좌표점에 node들이 발생할 수 있다. 발생된 node는 자신보다 하위 layer 값을 가지는 임의의 node들로부터

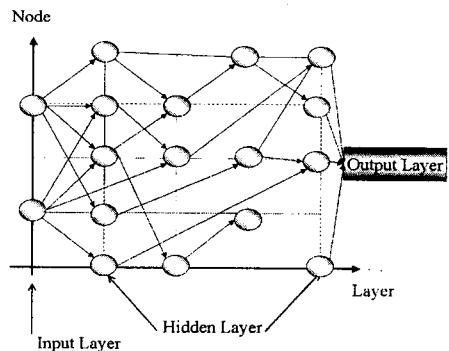


그림 2 신경회로망 발생 공간

Fig 2 Neural Network Space

터 입력 값을 받으면 자신보다 상위 layer값을 가지는 임의의 node들에게 출력을 내보낸다. 따라서 ENNet에서는 절대적인 layer의 의미는 없어지며 각 node의 연결에 따른 상대적인 layer가 구성된다. 동일 좌표 상에서도 여러 개의 node를 발생시킴으로써 Neural Network Space에서 발생할 수 있는 node수에 제한을 두지 않았다.

ENNet은 [-1,1]사이로 정규화된 입력 값을 가지며 [-1,1]사이의 출력 값을 내보낸다.

Layer축 i 번째, Node축 j 번째에 위치한 node는 Node<sub>(i,j)</sub>와 같이 나타내며 출력 식은 식(2)와 같다.

$$net_{(i,j)} = \sum_n W_{(i,j)(l,n)} i_{(l,n)} \quad (1)$$

$$i_{(i,j)} = \frac{2}{1 + \exp(-\alpha * net_{(i,j)})} - 1 \quad (2)$$

W<sub>(i,j)(l,n)</sub> : Weight from Node<sub>(l,n)</sub> to Node<sub>(i,j)</sub>

i<sub>(i,j)</sub> : Output of Node<sub>(i,j)</sub>

출력 node의 출력식은 (4)와 같다.

$$net_{(o,k)} = \sum_n W_{(i,j)(l,n)} i_{(l,n)} \quad (3)$$

$$O_{(o,k)} = \frac{2}{1 + \exp(-\alpha * net_{(o,k)})} - 1 \quad (4)$$

O<sub>(o,k)</sub> : Output of kth Outputnode<sub>(o,k)</sub>

입력 node는 (0, j)에 위치시킨다. 각 node의 bias는 고려하지 않았으며  $\alpha$ 는 상수로 주어진다.

#### IV. DNA 해석을 통한 ENNet 구성 방법

ENNet은 주어진 발생환경과 자신의 DNA에 포함된 유전정보를 통해서 발생된다. 발생환경은 입력의 개수, 출력의 개수, Neural Network Space의 크기 등으로 구성되며, 발생할 node들은 Neural Network Space상에서 자신의 위치 정보, 입력을 받을 수 있는 용량, 입력을 받을 node들의 위치와 weight등으로 구성된다.

하나의 chromosome에는 여러 개의 Gene을 포함하고 있다. 하나의 Gene은 신경회로망을 구성하는 하나의 node를 발생시키며 이러한 과정을 통해서 발생된 node들의 구성으로 전체 신경회로망이 구성된다. Gene의 구성은 그림 3과 같다.

DNA 해석과정은 두 단계로 구성된다. First Translation Phase에서는 node의 위치 정보와 연결 할

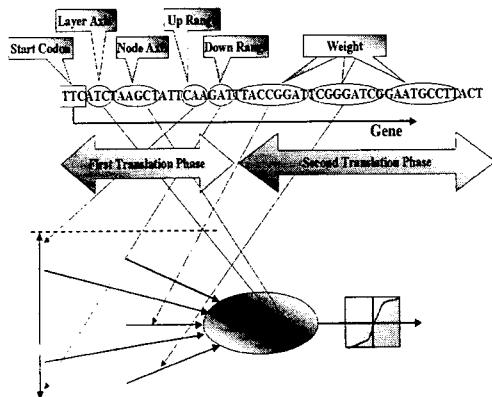


그림 3 Node 발생 모델과 Gene의 구조

Fig 3 Node Translation model & Gene structure

Table 1. Amino acid table

T(U)	C	A	G	T(E)
T T T	T C T	I A T	T G T	C
T T C	T C C	T A C	T G C	A
T T A	P k r	T C A	T G A	G
T T G	T r p	T C G	T G T	T(E)
C T T	C C I	C A T	C G T	C
C T C	C C C	C A C	C G C	C
C T A	L r n	C A A	C G A	A
C	C T G	G l n	C G G	G
A T T		A C T	A C T	T(E)
A T C		A C C	A G C	C
A T A	I l r	A C A	A G A	A
A	A T G	N e t	M e t	G
G T T	G C T	G A T	G C T	T(E)
G C C	G C C	G A C	G C C	C
G T A	V s t	G C A	G G C	A
G	G T G	G l y	G G G	G

Artificial Amino acid      Biological Amino acid

수 있는 용량정보가 해석되며 Second Translation Phase에서는 발생된 node사이의 연결상태와 weight정보가 해석된다.

Amino acid translation table을 Table 1과 같이 구성하였다. 일반적인 생물의 amino acid translation table과는 달리 대응하는 codon에서 amino acid들이 발생할 수 있는 확률을 최대한 균등하게 하며 비슷한 구성의 codon은 동일한 amino acid로 발생하도록 table을 구성하였다.

Table 1 을 바탕으로 First Translation Phase를 수행한다. start codon이 검출되면 그 다음 3개의 codon을 통해 Neural Network Space상에서 node가 차지할 위치정보를 변역한다. 그리고 그 다음 2개의 codon으로 발생한 node가 다른 node들과 연결할 수 있는 범위정보를 변역한다.

DNA상에서 발생한 node가 연결된 node들의 위치정보를 가지고 있지 않고 용량정보를 가짐으로써 그 node를 발생시킨 Gene이 다음 세대로 전파됐을 경우 위치정보가 바뀌더라도 node의 다른 특성을 보존되는 효과를 가져온다. 하나의 Gene에서 First Translation Phase가 수행되면 다음에 존재하는 Gene에 대해 First Translation Phase를 수행한다.

Chromosome내의 모든 Gene에 대해 First Translation Phase를 수행하면 그 다음은 존재하는 모든 Gene에 대해 Second Translation Phase를 수행한다. 발생된 node들은 자신의 Gene정보와 Neural Network Space상의 발생된 node정보로부터 link상태와 link의 weight 발생시킨다. 발생한 node가 가지고 있는 연결범위정보에 따라 다음에 해석될 codon의 수는 가변이다.

Amino acid들과 neural network의 형질사이의 대응관계는 Table 2와 같다.

Table 2. Amino acid translation table

< Codon Translation Table >																			
Pho	Ser	Ity	Cys	Irp	Len	Pro	His	Arg	Gln	Ile	Trp	Asn	Lys	Met	Val	Ala	Asp	Glu	Gly
Node Num	0	1	2	3	4	5	6	7	8	9									
Layer Num											1			2					
Range	1	2	3		4		5		6		7		8	9					
Weight Translation Table																			
Weight Translation Table																			
T(U)	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	
C	5	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
A	-1	0.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
G	-1	0.4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	

$$\text{Weight} = a^*b^*(c+d^*4^{-1}+e^*4^{-2}+f^*4^{-3}+g^*4^{-4}+h^*4^{-5}+i^*4^{-6})$$

## V. Simulation result

제시된 ENNet model를 여러 가지 함수 근사화 문제에 적용시켜 보았다.  
적합도 평가함수는 식 (6)과 같이 구성하였다.

$$Fit = \frac{1}{\alpha * error^2 + \beta * layer\# + \gamma * length} \quad (5)$$

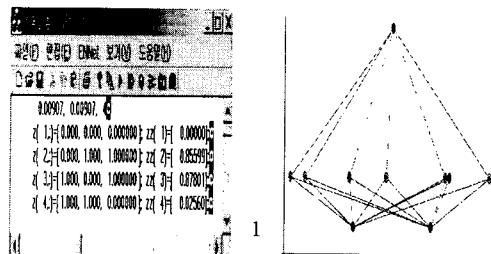
error는 학습 데이터에 대한 신경회로망 출력의 오차의 제곱 값을 사용하였으며 layer는 발생한 신경회로망의 layer수이고 length는 신경회로망을 발생하기 위한 DNA code의 길이로써 발생한 node개수에 비례한다.  $\alpha$ ,  $\beta$ ,  $\gamma$ 는 상수 값으로 주어진다. crossover는 one point cross over와 two point crossover를 병행해서 사용하였다.

초기 DNA길이는 500으로 설정하였다. 선택 방법은 elitist selection을 사용했다. Neural Network Space의 크기는 layer axis의 최대 값은 3 node axis의 최대 값은 30으로 설정했다

### i ) XOR Problem

Pop size : 50  
cross over rate : 0.8  
mutation rate : 3/(DNA length)  
 $\alpha : 1$ ,  $\beta : 0.3$ ,  $\gamma : 0.3$

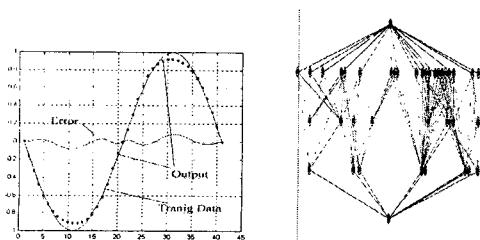
1000세대 후 구성된 신경회로망과 출력은 아래와 같다.



### ii ) Sine function approximation

Pop size : 100  
cross over rate : 0.5  
mutation rate : 10/(DNA length)  
 $\alpha : 1$ ,  $\beta : 0.0003$ ,  $\gamma : 0.0003$

500세대 이후 구성된 신경회로망과 출력은 아래와 같다

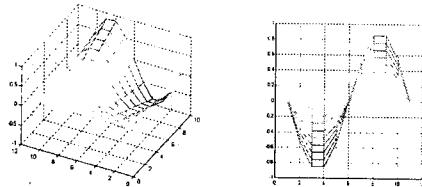


### iii) 2D function approximation ( $y = x_1 * \sin(x_2)$ )

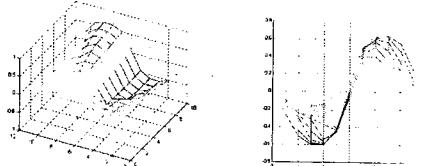
Pop size : 100  
cross over rate : 0.5  
mutation rate : 10/(DNA length)  
 $\alpha : 1$ ,  $\beta : 0.0003$ ,  $\gamma : 0.0003$

1000세대 후 구성된 출력은 아래와 같다.

#### <training data>



#### <Output>



#### 참고문헌

- [1] T. Yoshilawa, T. Furuhashi, Y.Uchidawa "Acquisition of Fuzzy Rules of Constructing Intelligent Systems using Genetic Algorithm based on DNA Coding Method" Proceedings of International Joint Conference of CFSA/IFIS/SOFT'95 on Fuzzy Theory and Applications
- [2] T. Yoshilawa, T. Furuhashi, Y.Uchidawa."The Effect of Combination of DNA Coding Method with Pseudo-Bacterial GA" Proceeding of the 1997 IEEE International Intermag. 97 Magnetics Conference, 4/01/1997
- [3] D. E. Goldberg "Genetic Algorithms in Search, Optimization & Machine Learning" Addison Wesley 1953
- [4] J. A. Freeman D. M. Skapura "Neural Network" Addison Wesley 1992