

**DNA 코딩 기법을 이용한 웨이브렛 기반 퍼지 모델링**

이 연우\* · 유 진영\* · 주 영훈\* · 박 진배\*\*  
 \*군산대학교 전자정보공학부, \*\*연세대학교 전기전자공학과

**Wavelet-Based Fuzzy Modeling Using a DNA Coding Method**

Yeun-Woo Lee\*, Jin-Young Yu\*, Young-Hoon Joo\*, Jin-Bae Park\*\*  
 \*School of Electronic & Information Eng, Kunsan National University,  
 \*\*Department of Electrical Electronic Eng, Yonsei University

**Abstract** - In this paper, we propose a new method about wavelet-based fuzzy modeling using a DNA coding method. DNA coding techniques is known that expression of knowledge is various than Genetic Algorithm(GA) usually by made optimization technique because done base in structure of biologic DNA and optimization performance is superior. The proposed method make fuzzy system model in wavelet transform and equivalence relation after identification with coefficient of wavelet transform using a DNA coding techniques. Also, can get fuzzy model effectively of nonlinear system using advantage of strong wavelet transform about function that have sudden change. In this paper, in order to demonstrate the superiority of the proposed method compared with GA.

한 결과와 기존 GA에 의한 방법과의 비교를 통해 그 우수성을 보인다.

**2. 이산 웨이브렛 변환 및 퍼지 모델**

**2.1 이산 웨이브렛 변환**

모든 함수 또는 신호  $f(x)$ 는 식(1)과 같이 기저함수들의 선형합으로 나타낼 수 있다.

$$f(x) = \sum_k c_{j_0, k} \varphi_{j_0, k}(x) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) \Psi_{j, k}(x) \quad (1)$$

여기서,  $\varphi_{j_0, k}(x)$ 는 스케일 함수,  $\Psi_{j, k}(x)$ 는 웨이브렛 함수라 한다. 식(1)에서 첫 번째의 합은 저해상도의 근사화이고 두 번째의 합은 고해상도의 근사화가 된다. 즉, 웨이브렛 변환은  $L^2(R) \rightarrow L^2(R^2)$ 의 매핑으로 저역 통과 필터인 스케일 함수와 고역 통과 필터인 웨이브렛 함수로 나눌 수 있다. 모든  $j, k$ 에 대해  $\varphi_{j_0, k}(x)$ 와  $\Psi_{j, k}(x)$ 가 서로 직교한다면 식 (1)의 계수  $c_j(k)$ 와  $d_j(k)$ 를 내적의 함으로 구할 수 있다. 여기서, 계수를 구하는 것을 웨이브렛 변환이라 하며 이러한 다해상도 수식은 신호를 자세한 성분들로 분해하여 나타냄으로써 신호를 시간-주파수로 표현할 수 있다. 따라서, 식 (1)의  $\Psi_{j, k}(x)$ 는 식 (2)로 나타낼 수 있다.

$$\Psi_{j, k}(x) = a^{-j/2} \Psi(a^{-j}x - bk) \quad (2)$$

또한, 다차원 웨이브렛 함수를 단일 차원의 웨이브렛 함수의 텐서곱(tensor product)으로 나타내면 식(3)과 같다.

$$\Psi(x) = \Psi_1(x_1) \cdots \Psi_n(x_n) \quad (3)$$

여기서, 단일 차원의 웨이브렛 변환은  $n$ 개의 직교방향으로 분리되어 있다고 가정하면 각각의 항이 푸리에 변환된 식은 다음과 같이 얻을 수 있다.

$$\hat{\Psi}(\omega) = \hat{\Psi}_1(\omega_1) \cdots \hat{\Psi}_n(\omega_n) \quad (4)$$

여기서,  $\hat{\Psi}(\omega)$ 는  $\Psi(x)$ 의 푸리에 변환이며 식 (5)의 허용(admissibility) 조건을 만족하여야 한다.

$$\int |\hat{\Psi}_i(\omega_i)|^2 / |\omega_i| d\omega_i < \infty \quad (5)$$

간재 조건하에서 식 (5)는 음과 양의 무한대에서 0으로 수렴하는 각각의  $\hat{\Psi}_i(x_i)$ 에 대해서 식 (6)과 같이 적분한 값이 0인 것을 선택한다면 식 (5)를 만족시킬 수 있다.

$$\int \Psi_i(x_i) dx_i = 0 \quad (6)$$

한편 웨이브렛 프레임이 되기 위해서는 식 (7)의 조건도 만족하여야 한다.

$$A \|f\|^2 \leq \sum_{j,k} |(f, \Psi_{j,k})|^2 \leq B \|f\|^2 \quad (7)$$

여기서,  $A > 0, 0 < B < \infty$ 이다.

따라서, 식 (6)과 식 (7)을 만족하는  $\Psi_i(x_i)$ 를 웨이브렛 프레임으

**1. 서 론**

복잡하고 불확실한 비선형 시스템을 수학적으로 모델링하는 경우에는 비선형성의 고차 모델을 이용하거나 모델 근사화 방법들을 이용하더라도 만족할 만한 모델을 얻기 어렵다. 이러한 시스템을 다루기 위해 *if-then* 규칙 형식으로 표현되는 퍼지 모델링 기법이 활발히 연구되어지고 있다. Zadeh(1)에 의해 최초로 도입된 퍼지 논리 개념은 각계에서 새로운 가능성을 보여주었다. Tong(2)은 퍼지 논리를 바탕으로 시스템을 묘사하는 퍼지 모델링 기법을 제안하여 복잡한 비선형 시스템의 모델링에 새로운 접근 방법을 제시하였다. 그후 Pedrycz(3), Xu(4) 그리고 Sugeno(5) 등에 의해 많은 모델링 기법들이 연구되었다.

상기 연구에서 알 수 있듯이 퍼지 모델링 방법의 주 관심사는 전문가적 경험을 이용하여 미지의 시스템을 퍼지 규칙으로 정확히 묘사하는 데 있다. 최근에는 전역 최적화 탐색 기법인 유전 알고리즘을 이용한 퍼지 모델링 방법의 많은 연구가 진행되고 있다. 그러나 유전 알고리즘은 비선형 성이 강한 시스템에 대하여 최적의 해를 찾지 못하고 국소적인 값에 빠질 위험이 있으며 최적의 해를 찾기 위해 많은 시간이 필요로 하게 된다. 이에 Joo(6, 7)는 유전 알고리즘과 클러스터링 기법을 이용해 퍼지 모델의 파라미터와 구조를 동정하는 자동 동정 기법과 유전 알고리즘의 변형된 형태의 mGA(messy Genetic Algorithm)를 이용한 퍼지 모델링 기법을 제안하였다. 또한 최근 이러한 유전 알고리즘의 문제점을 해결하기 위한 방법의 하나로 생물학적 DNA 구조를 이용한 DNA 코딩 기법이 연구되고 있다(8).

DNA 코딩 기법은 염색체의 가변적인 형태와 비트 단위가 아닌 생물의 DNA 구조에 의한 해석으로 유전 알고리즘보다 빠르게 최적 해를 찾을 수 있으며 다변적인 염색체의 구조와 지식의 융통성 있는 표현으로 최적 해를 구하는데 유전 알고리즘보다 유리하다고 알려져 있다(8). 또한 Zeng(9)은 퍼지 기저 함수(Fuzzy Basis Function, FBF)의 선형 합으로 퍼지 시스템을 표현하였고 Lin(10)은 퍼지 모델의 변형을 통하여 이산 웨이브렛 변환과의 동가 모델을 획득하였다. 이와 같이 현대의 디지털 신호처리와 영상 압축 및 기본적인 여러 알고리즘 수행의 속도 증진과 디지털 신호의 잡음을 제거하기 위해 웨이브렛이 많이 사용되고 있다.

본 논문에서는 DNA 코딩 기법을 이용한 웨이브렛 기반 퍼지 모델링에 대한 새로운 방법을 제안한다. 제안된 방법은 여러 개의 퍼지 규칙 기저를 갖으며 퍼지 모델의 변형을 통해서 웨이브렛 변형과 등가의 퍼지 모델을 형성, 웨이브렛 변형의 장점을 상속받게 된다. 또한 퍼지 시스템의 파라미터들과 규칙 수를 동정하기 위해 DNA 코딩기법이 사용된다. 제안된 방법은 대표적인 비선형 시스템의 퍼지 모델링에 적용

로 설정하면 된다. 본 논문에서는 위의 두 가지 필요조건을 만족시키는 mother 웨이브렛 함수로 식 (8)의 'Mexican Hat'을 이용한다. 이 함수는 가우시안 확률 밀도 함수를 두 번 미분한 함수에 비례한 함수에서 유도된다.

$$\Psi_j(x_i) = a_j(1 - a_j x_i^2)e^{-\frac{a_j x_i^2}{2}} \quad (8)$$

식 (8)을 식 (2)와 (3)에 대입하면 다음 식을 얻을 수 있다.

$$\Psi_{j,k} = a^{\frac{j}{2}} \alpha_1 [1 - \alpha_1 (a^{-j} x_1 - b_1 k_1)^2] e^{-\frac{\alpha_1 (a^{-j} x_1 - b_1 k_1)^2}{2}} \dots - a^{\frac{j}{2}} \alpha_{11} [1 - \alpha_n (a^{-j} x_n - b_n k_n)^2] e^{-\frac{\alpha_n (a^{-j} x_n - b_n k_n)^2}{2}} \quad (9)$$

## 2.2 퍼지 모델

퍼지 모델에서 변수들 사이의 관계는 퍼지 if-then 규칙들의 관계로 나타내며, 본 논문에서 사용된 규칙의 형태는 식(10)과 같다.

Rule  $i$ : If  $x_1$  is  $A_{i1}$ , ...,  $x_n$  is  $A_{in}$ ,  
Then  $y_i$  is  $d_i \alpha_1 (1 - \alpha_1 x_1^2) \alpha_2 (1 - \alpha_2 x_2^2) \dots \alpha_n (1 - \alpha_n x_n^2)$  (10)

Rule  $i$ 는  $i$ 번째 규칙,  $x_j$ 는  $j$ 번째 입력 변수,  $y_i$ 는  $i$ 번째 출력 변수,  $A_{ij}$ 는 가우시안 함수로 정의되는  $j$ 번째 입력의  $i$ 번째 규칙에 대한 소속 함수이며, 결론부는 식 (9)에서의 'Mexican Hat' 웨이브렛 함수의 가우시안 함수부분을 제외한 나머지 부분의 곱으로 이루어져 있다. 즉, 일반적인 퍼지 모델의 후반부를 웨이브렛 기반 퍼지 모델의 규칙을 나타내도록 함으로써 웨이브렛 변환과 등가의 모델을 획득 할 수 있다. 따라서 여러개의 퍼지 규칙 기저를 갖는 퍼지 추론 시스템에서 임의의 퍼지 규칙 기저에 대한 출력은 식 (11)와 같이 FBF의 선형 합으로 표현할 수 있다.

$$y_i = \sum_{j=1}^n B_j(x_j) y_j \quad (11)$$

식 (10)을 식 (11)에 적용하면 식 (12)을 얻을 수 있으며,

$$y_i = \sum_{j=1}^n B_j d_i a^{\frac{j}{2}} \alpha_1 (1 - \alpha_1 (a^{-j} x_1 - b_1 k_1)^2) \dots a^{\frac{j}{2}} \alpha_n (1 - \alpha_n (a^{-j} x_n - b_n k_n)^2) \quad (12)$$

여기서,  $B_j = \prod_{k=1}^i A_{jk}(x_k) / \sum_{k=1}^i \prod_{k=1}^i A_{jk}(x_k)$ 이며,  $d_i$ 는 실수이다. 따라서, 최종적인 출력  $y$ 는 식 (13)으로부터 구할 수 있다.

$$y = \sum_{i=1}^N y_i = \sum_{i=1}^N \sum_{j=1}^n d_{i,j} \Psi_{j,k} \quad (13)$$

식 (13)은 여러 개의 퍼지 규칙 기저를 갖는 퍼지 모델의 후반부를 변형함으로써 이산 웨이브렛 변환과 등가가 됨을 나타낸다. 즉, 'Mexican Hat' 웨이브렛 함수를 이용한 경우, 웨이브렛 함수내의 가우시안 함수부분이 퍼지 모델의 소속함수  $A_{ik}(x)$ 가 되고, 가우시안 함수부분을 제외한 부분의 곱이 식 (10)의 변형된 퍼지 모델의 후반부를 형성함으로써 변형된 퍼지 모델이 웨이브렛 변환과 등가임을 알 수 있다.

## 3. DNA 코딩 기법

DNA 코딩 기법은 생물학적 DNA의 구조에 기반 하여 만들어진 최적화 기법으로 일반적으로 유전 알고리즘보다 지식의 표현이 다양하고, 최적화 성능이 우수한 것으로 알려져 있다(8).

DNA는 하나의 세대에서 다른 세대로부터 전달되는 유전 정보의 첫 번째 전달자이며 DNA 분자들은 두 개의 길고 상보적인 체인들이 함께 엮여있는 기본적인 쌍들로 이루어져 있다. DNA는 sugar-phosphate를 중심으로 4종류의 염기들의 결합으로 이루어진다. DNA에서 4종류의 염기들은 A, G, T, C이며 염색체는 DNA 이중 나선으로 구성된다. DNA 나선들에서의 염기들은 상보적인 염기 접합 규칙을 따른다.

DNA에서 정보의 암호화는 유전자 합성의 과정동안 추출되어지고, 유전 정보 암호화 표현은 mRNA 배열의 구조를 필요로 한다. DNA는

RNA로 전사되어 리보솜에서 단백질로 번역된다. 즉 아미노산을 암호화하는 DNA의 배열에 따라 아미노산의 합성 순서를 결정하여 여러 종류의 단백질을 만들어 낸다. 그 중심 과정은 다음과 같다.

- transcription : formation of mRNA from DNA
- translation : formation of protein from mRNA
- protein folding

그림 1은 위의 과정을 나타낸 것이다.

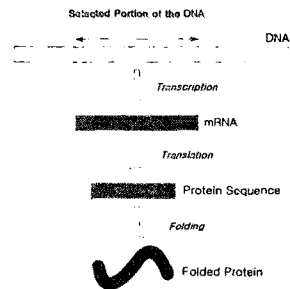
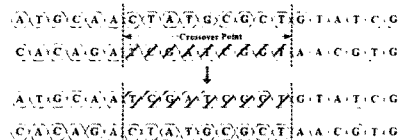
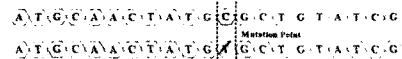


그림 1. 유전자 표현의 각 단계

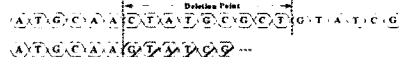
생물학적 DNA의 기본적 요소는 4개의 핵산(nucleotide)이다. 핵산은 A(Adenine), G(Guanine), C(Cytosine), 및 T(Thymine)으로 구성되고, 3개의 핵산이 결합하여 하나의 코돈(codon)을 형성한다. 이러한 64개의 코돈은 유전정보를 가지는 아미노산(amino acid)으로 표 1과 같이 번역되어진다. 표에서 각각의 아미노산이 가지는 유전정보를 나타내는 괄호 안의 값은 입력 및 출력 학습데이터 상에서 임의로 분할된 값을 지정하는 역할을 한다. 따라서 코돈은 아미노산으로 번역되고, 아미노산이 가지는 유전정보에 따라 퍼지규칙이 생성된다. 또한 각각의 개체들은 유전 연산자를 이용하여 진화하게 되며, 본 논문에서는 교차(Crossover), 돌연변이(Mutation), 삭제(Deletion), 삽입(Insertion) 및 도치(Inversion) 연산자를 사용하였고, 이러한 과정을 거쳐 적합도가 높은 개체는 다음 세대로 재생된다. 그림 2는 각각의 유전 연산자들의 연산 원리를 나타낸 것이다.



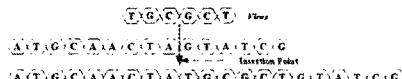
(a) Crossover Operator



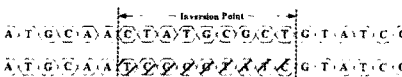
(b) Mutation Operator



(c) Deletion Operator



(d) Insertion Operator



(e) Inversion Operator

그림 2. 유전 연산자들의 연산원리

표 1. 코돈에서 유전정보를 갖는 아미노산 번역표

1st position (5' end)	2nd position(middle)				3rd position (3' end)
	T	C	A	G	
T	Phe(1)	Ser(10)	Tyr(4)	Cys(19)	T
	Phe(1)	Ser(10)	Tyr(4)	Cys(19)	C
	Leu(9)	Ser(10)	Stop	Stop	A
	Leu(9)	Ser(10)	Stop	Trp(20)	G
C	Leu(9)	Pro(7)	His(5)	Arg(11)	T
	Leu(9)	Pro(7)	His(5)	Arg(11)	C
	Leu(9)	Pro(7)	Gln(14)	Arg(11)	A
	Leu(9)	Pro(7)	Gln(14)	Arg(11)	G
A	Ile(2)	Thr(8)	Asn(15)	Ser(10)	T
	Ile(2)	Thr(8)	Asn(15)	Ser(10)	C
	Ile(2)	Thr(8)	Lys(16)	Arg(11)	A
	Met(3)	Thr(8)	Lys(16)	Arg(11)	G
G	Val(6)	Ala(12)	Asp(17)	Gly(13)	T
	Val(6)	Ala(12)	Asp(17)	Gly(13)	C
	Val(6)	Ala(12)	Glu(18)	Gly(13)	A
	Val(6)	Ala(12)	Glu(18)	Gly(13)	G

#### 4. 모의 실험

본 논문에서 제안된 퍼지 모델링 방법의 우수성을 입증하기 위해 Mackey-Glass 혼돈 시계열 시스템을 퍼지 모델링에 이용하였다. Mackey-Glass 혼돈 시스템은 식(14)과 같은 미분 지연 방정식으로 정의된다.

$$\dot{x}(t) = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \quad (14)$$

식 (15)는 시스템의 입출력 데이터 샘플을 추출하는데 사용된 식을 보여준다.

$$[x(t-18), x(t-12), x(t-6), x(t); x(t+6)] \quad (15)$$

식 (15)에서 혼돈 시계열 시스템의 퍼지 모델을 위해  $t=25$ 에서부터  $t=524$ 의 500개의 입출력 데이터 집합을 추출하였으며 제안된 방법의 초기 파라미터는 표 2와 같다.

표 2. 초기 파라미터

Parameter	Value
Generation (DNA, GA)	200
Population (DNA, GA)	300
Crossover rate (DNA, GA)	0.9
Mutation rate (DNA, GA)	0.01
Deletion rate (DNA)	
Insertion rate (DNA)	
Inversion rate (DNA)	

제안된 방법과 GA를 이용한 모의 실험 결과는 그림 3과 같다. 따라서 에너지가 몇몇 계수에 집중하는 웨이블릿 변환의 특성과 유전정보의 표현이 용이한 DNA 코딩 기법의 결합은 매우 효과적인 모델링 기법임을 알 수 있다. 표 3은 제안된 방법과 다른 알고리즘과의 NDEI를 비교한 것이다.

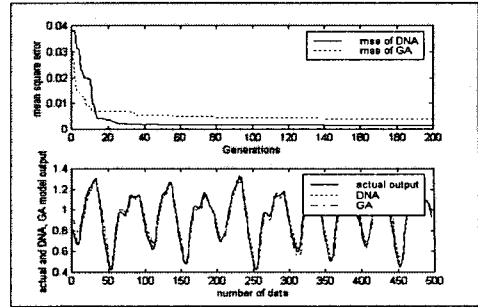


그림 3 시뮬레이션 결과

표 3. 제안된 방법과 다른 알고리즘과의 비교

Method	Non-dimensional error index
ANFIS	0.007
AR model	0.19
Cascaded-correlation NN	0.06
Backpropagation MLP	0.02
6th-order polynomial	0.04
Linear predictive method	0.55
Ours	0.001

#### 5. 결 론

본 논문에서는 DNA 코딩 기법을 이용한 웨이블릿 기반 퍼지 모델링 기법에 대한 새로운 방법을 제안하였다. 제안된 방법은 DNA 코딩 기법을 이용하여 웨이블릿 변환의 계수를 동정한 후 웨이블릿 변환과 등가관계에 있는 퍼지 시스템 모델을 형성, 퍼지 규칙의 구조와 파라미터를 동시에 동정시키는 모델링 기법이다. 본 논문에서는 기존의 GA에 비하여 좀 더 효과적인 암호화 기법인 DNA 코딩 기법과 급격한 변화를 갖는 함수에 대하여 강인한 웨이블릿 변환의 장점을 이용하여 대표적인 비선형 시스템인 Mackey-Glass 혼돈 시계열 시스템의 퍼지 모델링에 적용하였으며, GA와의 비교를 통하여 제안된 방법의 우수성을 입증하였다.

본 연구는 산업자원부의 지원에 의하여 기초전력공학공동연구구조주관으로 수행된 과제임(과제번호: R-2003-B-078).

#### [참 고 문 헌]

- [1] L. A. Zadeh, "Fuzzy Sets", *Int. Jour. of Information control*, vol. 8, pp. 338-353, 1965.
- [2] R. M. Tong, "The Evaluation of Fuzzy Models Derived from Experimental Data", *Fuzzy Sets and Systems*, vol. 4, pp. 1-12, 1980.
- [3] W. Pedrycz, "An Identification Algorithm in Fuzzy Relational Systems", *Fuzzy Sets and Systems*, vol. 13, pp. 153-167, 1984.
- [4] C. W. Xu, "Fuzzy Model Identification and Self-learning for Dynamic Systems", *IEEE Trans. Sys. Man, Cybern.*, vol. 17, No. 4, pp. 683-689, 1987.
- [5] M. Sugeno and T. Yasukawa, "A Fuzzy Logic Based Approach to Qualitative Modeling", *IEEE Trans. Fuzzy Sys.*, vol. 1, pp. 7-31, 1993.
- [6] Y. H. Joo, H. S. Hwang, K. B. Kim and K. B. Woo, "Fuzzy System Modeling by Fuzzy Partition and GA Hybrid Schemes", *Fuzzy Sets and Systems*, vol. 86, No. 3, pp. 279-288, 1997.
- [7] 추경훈, 이만우, 박근배, "매시 유전정보를 이용한 퍼지 모델 동정", *대한전자학회 논문지* vol. 49D, No. 8, pp. 407-414, 2000, 8.
- [8] T. Yoshikawa, T. Furuhashi, Y. Uchikawa, "A Combination of DNA Coding Method with Pseudo-Bacterial GA for Acquisition of Fuzzy Control Rules", *The 1st Online Workshop on Soft Computing(WCS)*, 1996, 8.
- [9] X. J. Zeng and M. G. Singh, "Approximation Accuracy Analysis of Fuzzy Systems with the Center-Average Defuzzifier", *Proc. FUZZ-IEEE/IFES'95*, pp. 109-116, 1995.
- [10] C. K. Lin and S. D. Wang, "Fuzzy Modeling Using Wavelet Transforms", *Electronics Letter*, vol. 32, pp. 2255-2256, 1996.