

## 유전자 알고리즘의 기호 코딩을 이용한 정보 입자기반 퍼지 다항식 뉴럴네트워크의 설계와 소프트웨어 공정으로의 응용

이인태, 오성권  
수원대학교 전기공학과

### Design of Information Granules based Fuzzy Polynomial Networks Using Symbolic Encoding of Genetic Algorithms and Its Application to Software Systems

In-Tae Lee, Sung-Kwon Oh  
Department of Electrical Engineering, The University of Suwon

**Abstract** - 본 논문은 소프트웨어 공정에 대하여 유전자 알고리즘의 기호 코딩을 이용한 정보입자 기반 퍼지 다항식 뉴럴 네트워크(Information Granules based genetic Fuzzy Polynomial Neural Networks; IG based gFPNN)의 모델 설계를 제안한다. 기존 퍼지 다항식 뉴럴네트워크의 구조 최적화를 위해 이진코딩을 사용하였다. 그러나 이진코딩에서 스트링의 길이가 길면 길수록 인접한 두 수 사이에 발생하는 급격한 비트 차이라는 해밍 절벽이 발생하였다. 이에 제안된 모델에서는 해밍절벽의 문제를 해결하기 위해 기호코딩을 사용하였다. 제안된 모델의 전반부 구조와 후반부 구조는 기존 모델에 구성을 그대로 사용한다. 실험적 예제를 통하여 제안된 모델의 근사화 능력과 일반화 능력이 우수함을 보인다.

#### 1. 서 론

지능형 기술에 대한 관심이 증가함에 따라 미국, 일본을 비롯한 세계 선진 국가들은 첨단 컴퓨터 지능 기술인 Computational Intelligence(CI) 기술 개발에 박차를 가하고 있다. CI는 약 10여년 전에 신경망, 입자화(Granular) 컴퓨팅, 그리고 진화 컴퓨팅(Evolutionary Computing)등의 매우 가시적이고 뛰어난 정보기술들의 컨소시엄(consortium:연합) 형태로써 나타났다. 최근에는 이러한 CI 기술을 위해 지능관련 컴퓨터 기술이 산업전반에 다양하게 활용되면서 이의 능력에 대해 많은 관심이 모아지고 있다. 최근엔 이러한 CI 기술에 데이터 특성을 효과적으로 반영하는 정보 granule(Information Granule; IG)[1,2]을 접목하여 새로운 형태의 고급 컴퓨터 지능형 기술이 개발되고 있다.[3]

이에 본 논문에 제안된 모델은 정보입자 기반 유전자적 퍼지 다항식 뉴럴네트워크(Information Granules based genetic Fuzzy Polynomial Neural Networks; IG based gFPNN)이다. 제안된 모델은 복잡성과 불확실성을 나타내는 비선형 시스템에서 다른 지능형 모델에 비해 보다 나은 특성을 가지고 있다. 이는 확장된 GMDH 알고리즘[4]과 퍼지 추론 시스템의 융합으로부터 얻어지는 네트워크로, 구조는 퍼지 다항식 뉴론(Fuzzy Polynomial Neuron; FPN)들의 결합으로 구성되어 있다. 비선형 시스템이 복잡해지면서 2입력 2차식만으로 구성된 GMDH 만으로는 한계가 있어 이를 다변수 다차식으로 확장한 다항식 뉴럴네트워크(Polynomial Neural Networks; PNN)와 PNN에 FIS(Fuzzy Inference System)를 융합한 퍼지 다항식 뉴럴네트워크(Fuzzy Polynomial Neural Networks; FPNN)[5,6]가 Oh등에 의해 개발되었다. 하지만 PNN과 FPNN의 경우 최적의 구조를 찾기 위해서는 많은 실험을 통하여 최적의 구조를 찾는 단점이 있었다. 특히 FPNN은 입력변수의 선택, 멤버쉽 개수의 선택, 후반부 다항식의 선택 등을 연구자가 직접 바꿔가며 최적의 구조를 찾아야 했고, 층의 진행에 따라 획일적인 구조로 최적의 구조를 찾는데는 문제점이 산재하였다. 이를 해결하는 방법으로 최적화 탐색 알고리즘인 유전자 알고리즘[7,8]을 사용하여 FPNN의 전반부와 후반부 구조를 최적화 시켰다. 여기에 IG를 접목시켜 FPN의 전반부 멤버쉽 함수의 정점을 HCM 클러스터링으로 결정함으로써 비선형 시스템을 보다 최적화하고, 후반부 다항식 역시 HCM 클러스터링 알고리즘에 의한 출력력 데이터간의 중심값을 이용하여 모델의 구축 속도를 빠르게 하였다.

본 논문의 평가를 위해 Medical Imagine System[9] 데이터를 이용하여 제안된 모델의 타당성, 정확성 및 예측 성능 등을 기준 지능 모델들과 비교를 통해 그 우수성을 보인다.

#### 2. Information Granules

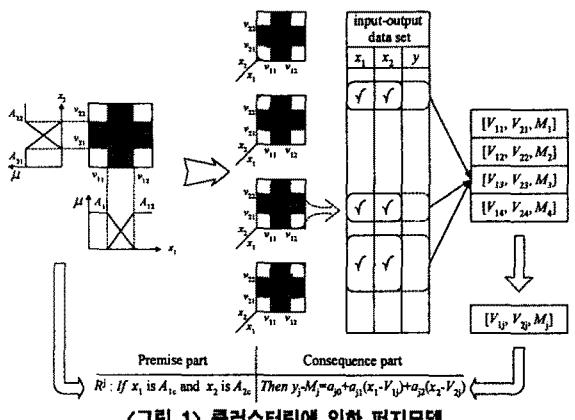
데이터 입자는 균질성, 유사성 또는 기능성 등의 기준에 의해 서로 결합된 물체(특히, 데이터 점)의 연결된 모임으로 간주되며, 기준보다 적은 계산적인 복잡성을 위해 잘 정의된 하위문제(모듈)들로 분할된다. 본 논문에서는 입출력 관계의 특성을 이용하여 모델링을 수행한다. 따라서 데이터의 특성을 파악하는 것이 매우 중요하다. 입출력 데이터의 특성을 찾아내기 위해 HCM 클러스터링 알고리즘을 이용한다.

본 논문에서는 데이터들간의 거리를 기준으로 균질한 정도를 측정하여 데이터를 특성별로 분류하는 HCM 클러스터링 알고리즘[4]을 이용한다. HCM 클러스터링에 의한 데이터 분류는 다음과 같다.

- [단계 1] 클러스터 개수 ( $2 \leq c \leq n$ )를 결정하고, 소속행렬  $U$ 를 초기화한다.
- [단계 2] 각각의 클러스터에 대한 중심벡터를 구한다.
- [단계 3] 각각의 클러스터 중심과 데이터와의 거리를 계산하여 새로운 소속행렬을 생성한다.

[단계 4] 만일 종료 조건을 만족한다면 종료하고, 그렇지 않으면 [단계 2]로 간다.

그림1은 HCM 클러스터링에 의한 2입력 1출력 시스템의 입출력 데이터셋을 가지고 퍼지 관계 기반 퍼지 모델(선형 퍼지추론)을 형성하는 과정을 보여준다.



〈그림 1〉 클러스터링에 의한 퍼지모델

#### 3. IG based gFPNN의 설계

데이터 입자화 기반 유전자적 퍼지 다항식 뉴럴 네트워크의 최적화 설계 과정은 다음과 같다.

- [단계 1] 출력 변수  $y$ 와 시스템 입력 변수 결정.
- [단계 2] 입출력 데이터를 학습용과 평가용 데이터로 분할한다.
- [단계 3] IG based gFPNN의 퍼지 추론 방법 및 동정을 설계하는 부분과 모델의 구조를 결정하는 초기 정보를 결정한다.
- [단계 4] FPNN 모델을 구축하는데 기본 유닛인 FPN을 유전자 알고리즘에 서 각각의 염색체에 의한 구조 선택을 통해 최적 모델 구축



〈그림 2〉 염색체구조

[단계 5] 전반부 동정

HCM 클러스터링을 통한 데이터 입자에 의한 전반부 동정은 다음과 같다. 전체 데이터 집합  $U=(x_1, x_2, \dots, x_l; y)$ 이고, 여기서  $x_k=[x_{1k}, \dots, x_{mk}]^T$ ,  $y=[y_1, \dots, y_n]^T$ ,  $l$ 은 변수의 수 그리고  $m$ 은 데이터의 수로 가정한다. 데이터 집합  $X_k$ 를  $c$ 개(멤버쉽 함수의 수)의 클러스터(데이터 입자)로 분류 한다.

$$v_{kg} = \{v_{k1}, v_{k2}, \dots, v_{kc}\} \quad (1)$$

여기서,  $k=1, 2, \dots, l$ ,  $g = 1, 2, \dots, c$ . 중심 벡터  $v_{kg}$ 를 멤버쉽함수의 정점으로 설정한다.

[단계 6] 후반부 동정

각 규칙에서 산술 평균에 의한 데이터 집합의 중심 벡터  $V_j$ 를 계산한다.

$$V_j = \{V_{j1}, V_{j2}, \dots, V_{jk}; M_j\} \quad (2)$$

여기서,  $k=1, 2, \dots, l$ ,  $j=1, 2, \dots, n$ .  $V_{jk}$ 와  $M_j$ 는 각각 입력 데이터와 출력 데이터의 중심값이다.

[단계 7] 모델 구축, 테스트 그리고 노드의 선택

[단계 8] [단계 7]의 현재 층에서 일어진 최적 노드의 최대 적합도 값  $F1$ 이 다음 부동식을 만족하는 경우에 알고리즘을 종료한다.

$$F1 \leq F*$$
 (3)

단,  $F*$ 은 이전 층의 최적 노드인 최대 적합도 값이다. 본 논문에서의 모델의 성능지수 식으로 다음 식을 이용하였다.

$$E = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (4)$$

[단계 9] 현재 층에서 보존된 노드의 출력들에서 다음 층의 새로운 입력들로,  $X_{lj}=Z_{lj}$ ,  $X_{2j}=Z_{2j}$ , ...,  $X_{Wj}=Z_{Wj}$ 으로 구성하고, [단계 4]로 간다(여기서,  $j=i+1$ ). 이 후, [단계 4]부터 [단계 9]까지를 반복한다.

#### 4. 실험 데이터를 통한 결과고찰

IG based gFPNN의 성능을 테스트하기 위해서 Medical Imagine System 데이터를 이용하였다. 표 2는 GAs정보와 FPNN 구축을 위한 정보를 나타낸 것이다.

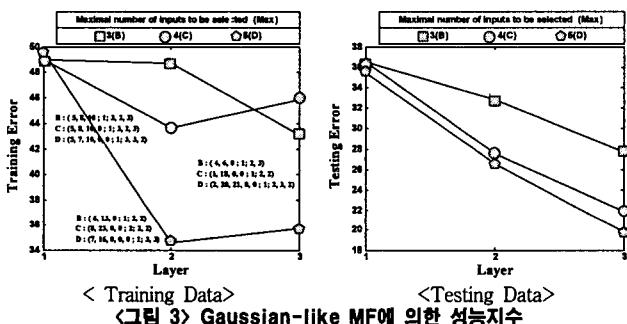
<표 2> GAs과 FPNN정보

	Parameters	1층~3층
GA	Maximum gen	300
	Total population size	150
	Selected population size	30
	Crossover rate	0.65
	Mutation rate	0.1
	String length	$1 + (\text{Max} * 2) + 1$
FPNN	Maximal no. of inputs to be selected (Max)	$\frac{1}{l} \leq l \leq \text{Max}(2\sim5)$
	polynomial Type(Type T) of the consequent part of rules	$1 \leq T \leq 4$
	Membership Function(MFs) type	Gaussian Triangular
	No. of MFs per each input	2 or 3

$l, T$  ; integer

Max ; 최대로 선택될 수 있는 입력변수의 수

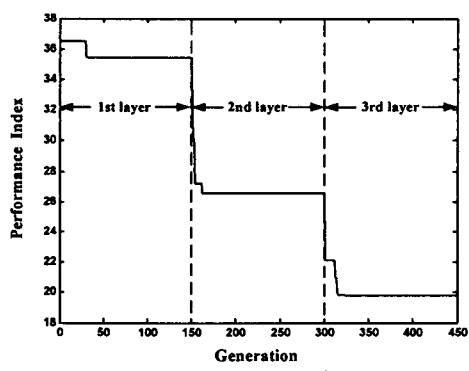
그림 2와 그림 3은 후반부 다항식에 선택입력을 사용했을 때의 최대 입력(Max)값에 따른 각 층별 성능지수를 나타내고 있다. 그림에서 B(4, 10, 0 ; 1; 3, 3)을 풀이하면 최대 입력이 3(B)일 때 선택된 입력변수는 4번과 10 번이고 후반부 다항식은 간략(1)을 선택하고 선택된 입력변수에 해당하는 멤버쉽 함수의 개수는 각각 3개를 나타내고 있다.



<그림 3> Gaussian-like MF에 의한 성능지수

그림에서 볼 수 있듯이 삼각형 멤버쉽 함수보다는 가우시안 멤버쉽 함수에서 성능이 우수함을 보여주고 있다. 또한 모델의 구조가 전체적으로 정해져서 동작하지 않고 그 데이터에 최적화되며 입력변수의 수와 후반부 다항식의 차수 그리고 각 입력에 대한 MF의 수가 유동적으로 변하여 최적의 모델을 탐색하는 것을 볼 수 있다.

그림 4는 가우시안 멤버쉽 함수이면서 Max=5일 때 generation 증가에 따른 성능지수를 보여주고 있다.



<그림 4> Generation 증가에 따른 성능지수

표 3은 기존의 FPNN과 제안된 모델의 성능지수를 비교한 표이다. 표에서 보는 바와 같이 제안된 모델이 전체적으로 성능이 우수함을 보여주고 있다. 특히 앞에서 보여준 그림의 출력인 Max=4이며, Gaussian-like 멤버쉽 함수일 때 최고의 성능 지수를 나타내고 있다. 기존의 FPNN 모델에 비해 3층에서 기존 모델과 거의 같은 성능을 보여주고 있다.

<표 2> 기존 모델과의 성능지수 비교

Model	Structure			PI	EPI
	Simplified	Generic Type	Basic architecture		
SONFN [10]	Linear	Generic Type	Basic architecture	Case 1	40.753
				Case 2	35.748
FPNN	Input	M_type	Layer	T	PI
				5	32.195
	3	G	5	1	49.716
				1	32.251
Our model	3	T	5	1	39.093
				2	30.798
	4	G	3	1	43.090
				2	39.816
5	T	3	4	1	45.870
				2	31.247
	G	3	1	1	35.706
				2	19.763

#### 5. 결론

본 논문에서는 IG-based gFPNN의 설계방법을 제안하였다. 기존의 gFPNN은 구조면에서 설계자의 경험 또는 학습이 아닌 객관적인 모델을 설계하기 위해 유전자 알고리즘을 사용하여 노드내의 입력변수의 수와 그에 따른 입력변수, 피지규칙 후반부 다항식의 차수 그리고 멤버쉽 함수의 수를 선택하여 모델을 설계하였다. 하지만 gFPNN은 데이터의 특성까지는 고려하지 않고 설계를 하기 위해 약간의 단점이 있었다. 이를 극복하고자 데이터 입자 기반 유전자적 최적 퍼지 다항식 뉴럴네트워크를 제안하였다. 이 모델은 gFPNN의 단점인 데이터의 특성까지도 고려한 모델로써 다변수 공정에서 보다 빠르게 최적 해를 찾을 뿐만 아니라 성능 또한 기존 gFPNN 보다 우수하였다. 또한 이진코딩에서 나타날 수 있는 해밍절벽의 문제를 해결하는 방안으로 기호코딩을 사용하여 보다 안정화하였다. 결과적으로 다변수 공정에서 IG-based gFPNN 모델이 우수한 예측 성능을 갖는다.

#### 감사의 글

본 연구는 산업자원부의 지원에 의하여 기초전력연구원(R-2003-B-274) 주관으로 수행된 과제임.

#### 참고 문헌

- L. A Zadeh, "Fuzzy logic = Computing with words," IEEE Trans. Fuzzy Syst., Vol. 4, No. 2, pp. 103-111, 1996.
- W. Pedrycz and G. Vukovich, "Granular neural networks," Neurocomputing, Vol. 36, pp. 205-224, 2001.
- Shinichi Horikawa, Takeshi Furuhashi and Yoshiki Uchigawa, "On Fuzzy Modeling Using Fuzzy Neural Networks with the Back propagation Algorithms", IEEE trans. Neural Networks, Vol.3, No.5, pp.801-806, 1992
- A. G. Ivakhnenko, "Polynomial theory of complex systems", IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-1, pp. 364-378, 1971.
- S.-K. Oh and W. Pedrycz, "The design of self-organizing Polynomial Neural Networks", Information Science, Vol. 141, pp. 237-258, 2002.
- S.-K. Oh and W. Pedrycz, "Fuzzy Polynomial Neuron-Based Self-Organizing Neural Networks", Int. J. of General Systems, Vol. 32, No. 3, p. 237-250, May, 2003.
- Holland, J. H. Adaptation In Natural and Artificial Systems, The University of Michigan Press, Ann Arbor, 1975.
- K. De Jong, Are genetic algorithms function optimizers? In Proc. of PPSN II (Parallel Problem Solving from Nature), pages 3-13, Amsterdam, North Holland, 1992.
- D. E. Box and G. M. Jenkins, Time Series Analysis, Forecasting and Control, California: Holden Day, 1976.
- S. K. Oh, W. Pedrycz, and B. J. Park, "Relation-based Neurofuzzy Networks with Evolutionary Data Granulation", Methematical and Computer Modeling, 2003.