

진화 알고리즘 기반 FPNN의 최적 동정 및 비선형 데이터로의 응용

Optimized Identification of Genetic Algorithms based FPNN and Its Application to Nonlinear Data

이인태*, 이동윤**, 김현기*, 오성권*

*수원대학교 전기공학과

**중부대학교 정보통신공학부

In-Tae Lee, Dong-Yoon Lee, Hyun-Ki Kim, Sung-Kwun Oh

*Dept. of Electrical Engineering, The University of Suwon

E-mail : ohsk@suwon.ac.kr

**Dept. of Information and Communication Engineering, Joongbu Univ.

요 약

본 논문은 유전자 알고리즘 기반 퍼지 다항식 뉴럴네트워크(Genetic Algorithm-based Fuzzy Polynomial Neural Networks ; GAS-based FPNN)를 이용하여 비선형 데이터의 최적화 추론 알고리즘을 제안한다. FPNN의 각 노드는 GMDH와 퍼지규칙을 기초로 만들었다. FPNN의 각 노드는 퍼지 다항식 뉴런(Fuzzy Polynomial Neuron ; FPN)이라고 표현하다. 제안된 모델은 구조 선택에 있어서 유전자 알고리즘(Genetic Algorithms ; GAs)을 이용하였다. 유전자 알고리즘을 사용하여 입력의 차수와 입력의 개수 그리고 후반부 추론의 형태를 최적 선택하였다. 비선형 데이터에 대한 모델 설계를 위해 최적화 알고리즘인 유전자 알고리즘 기반 FPNN 모델 설계가 유용하고 효과적임을 보인다.

1. 서론

대부분의 자연현상들은 근본적으로 비선형을 내포하고 있기 때문에 선형성을 바탕으로 개발된 기존의 방법들을 이용해 이들을 해석하기에는 많은 어려움에 부딪히게 된다. 초기에는 비선형성을 근사적으로 선형화 시킴으로써 이를 극복하려는 시도가 이루어졌으나 근본적인 한계를 넘지는 못하였다. 현재까지 개발된 모델들은 입력변수가 증가하면 시스템을 모델링하는데 어려움을 겪는다. 또한 시스템이 복잡하고 대규모적이면 함수식으로 표현하는데 한계가 있기 때문에 모델링을 설계하는데 회귀 분석을 적용할 수 없는 경우도 있다.

이러한 문제를 풀기 위해, Oh 등에 의해 자기구성 다항식 뉴럴 네트워크(Self-Organizing Polynomial Neural Networks ; SOPNN) [2]와 자기구성 퍼지 다항식 뉴럴 네트워크

(Self-Organizing Fuzzy Polynomial Neural Networks ; SOFPNN) [3]가 제안되었다. SOFPNN (또는 FPNN)의 구조는 퍼지 추론 규칙을 기반으로 한 활성화노드 즉, 퍼지 다항식 뉴런(Fuzzy Polynomial Neuron ; FPN)들로 구성되어 있으며, 일반적인 다층 퍼셉트론 구조의 경우와 달리 미리 정해져 있지 않고 층이 진행되는 동안 모델 스스로 노드의 선택과 제거를 통해 최적의 네트워크 구조를 생성할 수 있는 유연성을 가지고 있다. 그러나 기존 FPNN 구조는 성능을 향상시키기 위하여, 노드의 입력변수의 수와 규칙 후반부 다항식 차수 그리고 입력변수를 설계자의 경험 또는 반복적인 학습을 통해 선택된 네트워크 구조를 선택하였으나, 최적의 네트워크 구조를 구축하는데는 어려움이 내재되어 있었다.

본 논문에서는 FPNN을 최적화시키기 위해서

최적화 탐색 방법인 유전자 알고리즘[4, 5]을 이용하여 FPNN의 입력변수의 수와 이에 해당되는 입력변수 그리고 규칙 후반부 다항식의 차수를 탐색하여 최적의 퍼지 다항식 뉴럴 네트워크를 구축한다.

본 논문의 평가를 위해서 소프트웨어 공정 데이터인 Medical Imaging System(MIS) 데이터 [6]를 이용하여 제안된 모델의 타당성, 정확성 및 예측 성능 등을 기존 지능 모델들과 비교를 통해 그 우수성을 보인다.

2. Fuzzy Polynomial Neuron (FPN)

FPN은 FPNN의 가장 기본이 되는 처리 단위이다. FPN은 퍼지 집합과 뉴럴 네트워크의 개념에 기초해서 자세히 설명된다. 그림 1과 같이 2개의 기본적인 기능 모듈로 구성되어 있다.

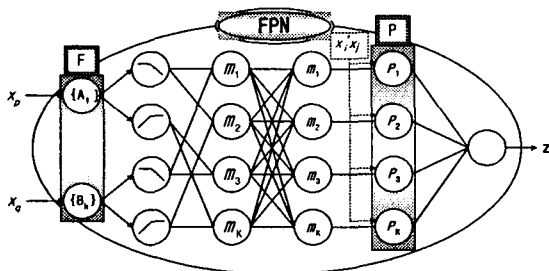


그림 1 FPN 모듈의 일반적인 구조

그림 1은 입력변수가 2개이고 멤버십 함수가 2x2인 경우를 보여주고 있다. "F"로 표기되어진 것은 퍼지 집합의 집합체로써 입력 수의 변수들과 뉴론에 의해 실현된 처리 부분사이를 상호 연결해 준다. 그림에서 Xp와 Xq를 입력변수들로 표시한다. "P"로 표기된 두 번째 모듈은 함수 기반 비선형(다항식) 처리에 관한 것이다.

3. GAs-based FPNN 설계

유전자 알고리즘을 이용한 퍼지 다항식 뉴럴 네트워크의 최적화 설계과정은 다음과 같다.

- [단계 1] 출력변수 y에 관계하는 n개의 시스템 입력변수를 결정한다.
- [단계 2] N개의 입출력 데이터를 학습용 데이터와 평가용 데이터로 분할한다.
- [단계 3] GAs-based FPNN의 퍼지 추론 방법 및 동정을 설계하는 부분과 모델의 구조를 결정하는 초기 정보를 결정한다.
- [단계 4] GAs 기반 FPNN 모델을 구축하는데 기본 유닛인 FPN 노드들을 유전자 알고리즘에서 각각의 염색체에 의한 구조 선택을 통해 최적

모델을 구축한다.

- [단계 4-1] 입력변수의 수 선택
 - [단계 4-2] 퍼지규칙 후반부 다항식 차수 선택
 - [단계 4-3] 입력변수 선택
- [단계 4-1]부터 [단계 4-3]까지 수행해서 얻은 선택된 입력변수와 퍼지규칙 후반부 다항식 차수를 이용하여 퍼지 추론을 적용한다.

표 1 퍼지규칙 후반부 구조의 서로 다른 형태

입력수 차수	1	2	3
0(Type 1)	Constant		
1(Type 2)	Linear	Bilinear	Trilinear
2(Type 3)		Biquadratic	Triquadratic
2(Type 4)	Quadratic	Modified Biquadratic	Modified Triquadratic

퍼지 추론 규칙과 파라미터 동정을 위한 매개 변수 추정은 다음과 같다.

i) 간략 추론

후반부가 단일의 상수항만을 가지는 것으로 이와 같은 추론법을 간략 추론법이라 한다.

$$R^j : \text{If } x_1 \text{ is } A_{j1} \text{ and } \dots \text{ and } x_n \text{ is } A_{jn} \text{ then } y_j \text{ is } a_{j0} \quad (1)$$

i번째 데이터의 추론된 값은 다음과 같다.

$$\hat{y}_i = \frac{\sum_{j=1}^z w_{ji} y_j}{\sum_{j=1}^z w_{ji}} = \frac{\sum_{j=1}^z w_{ji} a_{j0}}{\sum_{j=1}^z w_{ji}} = \sum_{j=1}^z \hat{w}_{ji} a_{j0} \quad (2)$$

$$w_{ji} = A_{j1}(x_{1i}) \wedge \dots \wedge A_{jn}(x_{ni})$$

여기서, i는 i(i=1, ..., m)번째 데이터, m은 데이터 총 수이다. 후반부 파라미터 a_{j0}의 동정은 입출력 데이터가 주어졌을 때 최소자승법에 의해 구해진다.

ii) 회귀다항식 추론

후반부가 1, 2, 변형된 2차 다항식 형태로 표현된 것으로 회귀다항식 추론이라 한다. 이 경우의 퍼지 모델은 다음과 같은 형태를 가지는 규칙들로 구성된다.

$$R^j : \text{If } x_1 \text{ is } A_{j1} \text{ and } \dots \text{ and } x_n \text{ is } A_{jn} \text{ then } y_j = f_j(x_1, \dots, x_n) \quad (3)$$

[단계 5] 모델 구축과 테스트 그리고 노드의 선택

- Step 1) GAs 기반의 FPNN 모델 구축을 위한 유전자의 세대, Population 개수, 돌연변이율, 교배율, 그리고 염색체의 길이와 같은 유전자의 정보를 설정한다.
- Step 2) 테스트 데이터를 이용해서 구해진 성능 지수를 유전자 알고리즘의 적합도 함수로 계산하

기 위해서 다음 식 (4)와 같이 변환하여 유전자 알고리즘의 적합도 함수로 사용한다.

$$F(\text{Fitness Function}) = 1/(1+EPI) \quad (4)$$

Step 3) 다음 세대의 생성을 위해서 구해진 개체군 정보와 적합도 값들을 이용하여 선택연산, 교배, 돌연변이를 실행한다.

Step 4) 구해진 적합도 값들(F_1, F_2, \dots, F_z)을 선호되는 적합도 값 순(내림차순)으로 정렬한다.

Step 5) 정렬된 적합도 값들 중에서 중복된 적합도 값들은 제거한다.

Step 6) 중복된 적합도 값들을 하나로 처리한 후 설계자가 FPNN 구조를 설계하기전 결정하는 노드의 제한된 개수 W개만큼 선호되는 적합도 값들을 선택한다.

Step 7) Step 6)에서 선택되어진 노드들을 다시 초기 population 번호 순서에 따른 오름차순으로 재정렬한다.

Step 8) 다음세대에서의 엘리트 전략을 위해서 선택된 노드들 중 최적의 적합도 값을 가지는 노드를 선택한다.

Step 9) Step 3)에서 얻어진 정보를 가지고 다음 세대의 population을 생성한 후 Step 2)에서 Step 8)까지 반복한다. 여기서 Step 2)에서 생성된 적합도 값들 중에서 가장 낮은 적합도 값을 가진 노드를 Step 8)에서 구한 이전 세대의 최적의 적합도 값을 가진 노드로 대치하는 엘리트 전략을 실행한다.

Step 10) 이전 세대에서 구해진 적합도 값들과 현재 세대에서 구해진 적합도 값들을 결합하여 다시 Step 4)부터 Step 7)까지 반복하고 정해진 세대까지 반복 실행한다.

[단계 6] [단계 5]의 현재 층에서 얻어진 최적 노드의 최대 적합도 값 F_1 이 다음 부등식을 만족하는 경우에 알고리즘을 종료한다.

$$F_1 \leq F^* \quad (5)$$

단, F^* 는 이전 층의 최적 노드인 최대 적합도 값이다. 본 논문에서의 모델의 성능지수식으로 다음 식을 이용하였다.

$$E = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (6)$$

[단계 7] 현재 층에서 보존된 노드의 출력들($z_{1i}, z_{2i}, \dots, z_{wi}$)에서 다음 층의 새로운 입력들($x_{1j}, x_{2j}, \dots, x_{wj}$)로, $x_{1j}=z_{1i}, x_{2j}=z_{2i}, \dots, x_{wj}=z_{wi}$ 으로 구성하고, [단계 4]로 간다(여기서, $j=i+1$). 이후, [단계 4]부터 [단계 7]까지를 반복한다. 알고리즘이 종료할 경우, 마지막 층의 최적 적합도 값 F^* 을 얻은 노드의 다항식에 입력이 되는 전 층의 출력을 대입하고 동일한 조작을 제 5층까지 반복하여 최종 추정 모델 \hat{y} 를 얻는다.

4. 실험 데이터를 통한 결과 고찰

제안된 모델의 성능을 테스트하기 위해서 MIS 데이터 [6]를 이용하였다. 표 2는 유전자 탐색체 정보와 퍼지 다항식 뉴럴 네트워크 구축을 위한 정보를 나타낸 것이다.

표 2 GA와 FPNN정보

Layer	Parameters	1층~5층
GA	Maximum gen	100
	Total population size	60
	Selected population size	30
	Crossover rate	0.65
	Mutation rate	0.1
	String length	3+3+30
FPNN	Maximal no. of inputs to be selected(Max)	$1 \leq l \leq \text{Max}(2 \sim 5)$
	polynomial Type(Type T) of the consequent part of rules	$1 \leq T \leq 4$
	Membership Function(MFs) type	1.Gaussian 2.Triangular
	No. of MFs per each input	2

l, T ; interger

* ; 후반부 다항식 입력이 전체 시스템 입력으로 사용된 경우

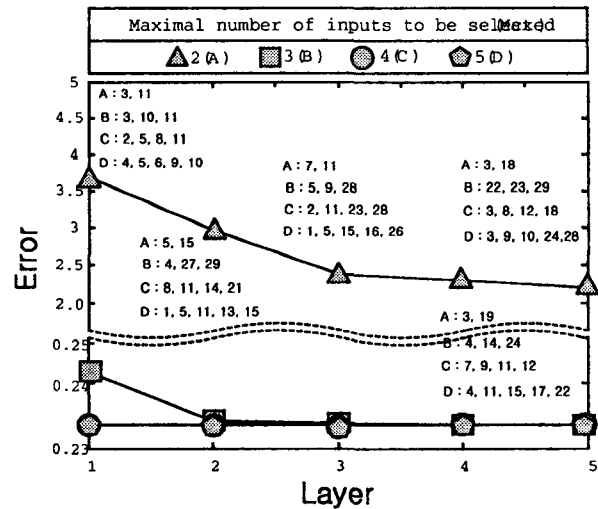


그림 2 Triangular MFs일때 성능지수

그림 2와 그림 3은 선택된 입력의 변화에 따른 GAs-based FPNN 각 층의 성능지수를 나타낸 것이다. 그림 4는 입력변수의 수가 2입력일때 GAs-based FPNN의 기본이 되는 처리 단위인 FPN의 경쟁식 결합과정을 보여주는 그림이다. 이처럼 GAs-FPNN은 기존의 모델보다 입력변수와 입력변수의 수 선택에 있어서 최적화 되는 과정이 FPNN처럼 고정되어서 결합되어지는 것이 아니라 유연하게 처리되는 것을 볼 수 있다.

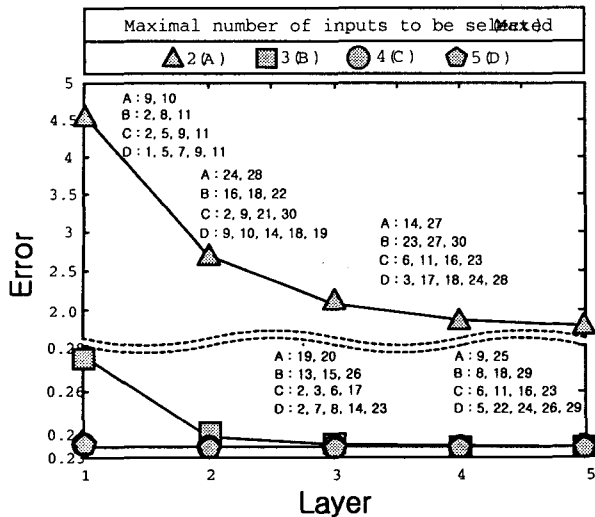


그림 3 Gaussian MFs일때 성능지수

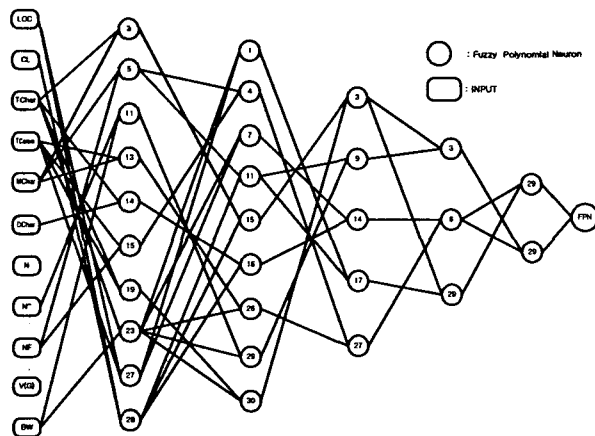


그림 4 GAs-based FPNN의 네트워크 구조

표 3 기존의 모델과 동정오차비교

	Max	M_type	T	PI
FPNN [3]	3	T	3	14.2660
		G	3	6.0077
	4	T	2	8.3247
		G	2	5.4009
Our model	2	T	3	2.1918
		G	3	1.7658
	3	T	4	0.23467
		G	3	0.23483
	4	T	2	0.23462
		G	4	0.23462
	5	T	2	0.23462
		G	4	0.23462

T : Triangular , G : Gaussian-like

표 3은 MIS데이터[6]를 사용하여 성능지수 식 (6)에 의해 계산한 값이다. Max=4, 5일때 그리고 후반부 구조가 삼각형 멤버쉽일때는 선형추론을 가우시안형일때는 변형된 회귀다항식 추론을 사용했을 때 PI성능이 우수함을 보인다. 또한 기존의 FPNN보다 Our model이 Max값과

M_type의 형태 그리고 차수(T) 선택에 있어서 좀 더 유연성과 정확성을 가지고 예측능력이 우수함을 보인다.

5. 결론

본 논문에서는 GAs-based FPNN의 설계방법을 제안하였다. 기존의 FPNN 구조는 성능을 향상시키는데 영향을 미치는 네트워크 층, 노드내의 입력변수의 수와 퍼지규칙 후반부 다항식의 차수 그리고 입력변수 선택을 설계자의 경험에 의해서 결정함으로써 객관적이고 주어진 데이터의 특성에 맞는 최적의 모델을 구축하는데 어려움이 있었다. 이러한 FPNN 모델 구조의 한계를 극복하고자 본 연구에서는 최적화 방법인 유전자 알고리즘을 이용하여 노드의 입력변수의 수와 그에 해당하는 입력변수 그리고 후반부 다항식 차수 등의 파라미터를 선택하였다. 결과적으로 모델 구축에 있어서 유연성과 정확성을 가지며 더욱 정확한 예측 능력을 가진 최적화된 GAs-based FPNN 모델 구조를 구축할 수가 있었다.

감사의글

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

6. 참고문헌

[1] A. G. Ivakhnenko, "Polynomial theory of complex systems", *IEEE Trans. on Systems, Man and Cybernetics*, Vol. SMC-1, pp. 364-378, 1971.

[2] S.-K. Oh and W. Pedrycz, "The design of self-organizing Polynomial Neural Networks", *Information Science*, Vol. 141, pp. 237-258, 2002.

[3] S.-K. Oh and W. Pedrycz, "Fuzzy Polynomial Neuron-Based Self-Organizing Neural Networks", *Int. J. of General Systems*, Vol. 32, No. 3, pp. 237-250, May, 2003.

[4] Holland, J. H., *Adaptation In Natural and Artificial Systems*, The University of Michigan Press, Ann Arbour. 1975.

[5] 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.

[6] Michael R. Lyu, "Handbook of Software Reliability Engineering", *Computing McGraw-Hill*, pp. 510-514, 1996