

유전론적 최적 퍼지 다항식 뉴럴네트워크와 다변수 소프트웨어 공정으로의 응용

Genetically Optimized Fuzzy Polynomial Neural Networks and Its Application to Multi-variable Software Process

이인태*, 오성권**, 김현기***, 이동윤****

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

Abstract - In this paper, we propose a new architecture of Fuzzy Polynomial Neural Networks(FPNN) by means of genetically optimized Fuzzy Polynomial Neuron(FPN) and discuss its comprehensive design methodology involving mechanisms of genetic optimization, especially Genetic Algorithms(GAs). The design of the network exploits the extended Group Method of Data Handling(GMDH) with some essential parameters of the network being provided by the designer and kept fixed throughout the overall development process. This restriction may hamper a possibility of producing an optimal architecture of the model. The proposed FPNN gives rise to a structurally optimized network and comes with a substantial level of flexibility in comparison to the one we encounter in conventional FPNNs. It is shown that the proposed genetic algorithms-based Fuzzy Polynomial Neural Networks is more useful and effective than the existing models for nonlinear process. We experimented with Medical Imaging System(MIS) dataset to evaluate the performance of the proposed model.

Key Words : Fuzzy Polynomial Neural Networks(FPNN), Genetic Algorithms(GAs), Fuzzy Polynomial Neuron(FPN), Group Method of Data Handling(GMDH), Software Process

1. 서론

대부분의 시스템 공정들은 모델의 차원이 증가하면 시스템을 모델링 하는데 어려움을 겪는다. 현재까지 개발된 모델들은 입력변수가 증가하면 시스템을 모델링 하는데 어려움이 있다. 또한 시스템이 복잡하고 대규모적이면 함수식으로 표현하는데 한계가 있기 때문에 모델링을 설계하는데 회귀 분석을 적용할 수 없는 경우도 있다.

이러한 문제를 풀기 위해, A. G. Ivakhnenko는 Group Method of Data Handling(GMDH) 알고리즘을 소개하였다.[1] 그러나 GMDH는 비교적 간단한 시스템에 대해서 매우 복잡한 다항식을 생성하는 경우가 있으며, 2입력, 2차 회귀다항식의 형태로 복잡한 비선형 시스템을 표현하는 데는 어느 정도 한계가 있다. GMDH의 이러한 단점을 극복하고자, 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 일반적인 다층 퍼셉트론 구조의 경우와 달리 미리 정해져 있지 않고 층이 진행되는 동안 모델 스스로 노드의 선택과 제거를 통해 최적의 네트워크 구조를 생성할 수 있다. 그러나 기존 FPNN 구조는 성능을 향상시키

기 위하여, 노드의 입력변수의 수와 규칙 후반부 다항식 차수 입력변수 그리고 멤버십 함수의 수를 설계자의 경험 또는 반복적인 학습을 통해 네트워크 구조를 선택하였으나, 최적의 네트워크 구조를 구축하는 데는 어려움이 내재되어 있었다.

본 논문에서는 FPNN을 최적화시키기 위해서 최적화 탐색 방법인 유전자 알고리즘[4, 5]을 이용하여 FPNN의 입력변수 수와 입력변수, 규칙 후반부 다항식의 차수 그리고 멤버십 함수의 수를 탐색하여 최적의 퍼지 다항식 뉴럴 네트워크를 구축한다. 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. Advanced GAs-based FPNN 설계

이 장에서는 유전자 알고리즘을 이용한 퍼지 다항식 뉴럴 네트워크의 최적화 설계과정을 설명한다.

[단계 1] 출력변수 y 에 관계하는 n 개의 시스템 입력변수를 결정한다. N 개의 입출력 데이터를 학습용 데이터와 평가용 데이터로 분할한다.

[단계 2] GAs 기반 FPNN 구조를 구축하는데 있어서 기본적인 정보를 결정한다.

[단계 3] GAs 기반 FPNN 모델을 구축하는데 기본 유닛인 FPN 노드들을 유전자 알고리즘에서 각각의 염색체에 의한 구조 선택을 통해 최적 모델을 구축할 수가 있다.

[단계 3-1] 입력변수의 수 선택

[단계 3-2] 퍼지규칙 후반부 다항식 차수 선택

* 準會員 : 水原大學校 電氣工學科 碩士課程

** 正會員 : 水原大學校 電氣工學科 教授

*** 正會員 : 水原大學校 電氣工學科 教授

**** 正會員 : 中部大學 情報通信工學部 助教授

[단계 3-3] 입력변수 선택

[단계 3-4] 멤버십 함수의 수 선택

[단계 3-1]부터 [단계 3-4]까지 수행해서 얻은 선택된 입력변수와 퍼지규칙 후반부 다항식 차수를 이용하여 퍼지 추론을 적용한다.

표 1. 퍼지규칙 후반부에서 회귀다항식의 서로 다른 형태

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

[단계4] 모델의 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) 첫 번째 세대에서 설정된 population 개수만큼 염색체에 의한 노드(FPN)을 구축한다.

Step 3) 테스트 데이터를 이용해서 구해진 성능지수를 유전자 알고리즘의 적합도 함수로 계산하기 위해서 다음 식(1)과 같이 변환하여 유전자 알고리즘의 적합도 함수로 사용한다.

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

Step 4) 다음 세대의 생성을 위해서 구해진 개체군 정보와 적합도 값들을 이용하여 선택연산, 교배, 돌연변이를 실행한다.

Step 5) 구해진 적합도 값들(F_1, F_2, \dots, F_z)을 선호되는 적합도 값 순(내림차순)으로 정렬한다.

Step 6) 정렬된 적합도 값들 중에서 중복된 적합도 값들은 제거하고 하나의 적합도 값으로 처리한다.

Step 7) 중복된 적합도 값들을 하나로 처리한 후 설계자가 FPNN 구조를 설계하기 전 결정된 노드의 제한된 개수 W개 만큼 선호되는 적합도 값들을 선택한다.

Step 8) Step 7)에서 선택되어진 노드들을 다시 초기 population 번호 순서에 따른 오름차순으로 재정렬한다.

Step 9) 다음세대에서의 엘리트 전략을 위해서 선택된 노드들 중에서 최적의 적합도 값을 가지는 노드를 선택한다.

Step 10) Step 4)에서 얻어진 정보를 가지고 다음 세대의 population을 생성한 후 Step 2)에서 Step 9)까지 반복한다.

Step 11) 이전 세대에서 구해진 적합도 값들과 현재 세대에서 구해진 적합도 값들을 결합하여 다시 Step 5)부터 Step 8)까지 반복한다.

Step 12) 정해진 세대까지 Step 10)부터 Step 11)을 반복 실행한다.

[단계 5] [단계 4]의 현재 층에서 얻어진 최적 노드의 최대 적합도 값 F_1 이 다음 부등식을 만족하는 경우에 알고리즘을 종료한다.

$$F_1 \leq F. \quad (2)$$

단, F,는 이전 층의 최적 노드인 최대 적합도 값이다. 본 논문에서의 모델의 성능지수를 구하는 식으로 다음 식을 이용하였다.

$$E = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (3)$$

[단계 6] 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 3]으로 간다(여기서, $j=i+1$). 이 후, [단계 3]부터 [단계 6]까지를 반복한다. 알고리즘이 종료할 경우, 마지막 층의 최적 적합도 값 F,을 얻은 노드의 다항식에 입력이 되는 전 층의 출력을 대입하고 동일한 조작을 세 5층까지 반복하여 최종 추정 모델 \hat{y} 를 얻는다.

3. 실험 데이터를 통한 결과 고찰

Advanced GAs-based FPNN의 성능을 테스트하기 위해서 Medical Imaging System 데이터[6]를 이용하였다. 이 공정은 기존의 이용할 수 있는 데이터를 보다 풍부한 다변수 데이터들로 구성되어 있기에 모델을 평가하는데 많은 도움이 된다. 표 2는 각각의 FPNN마다 GAs정보와 퍼지 다항식 뉴럴 네트워크 구축을 위한 정보를 나타낸 것이다.

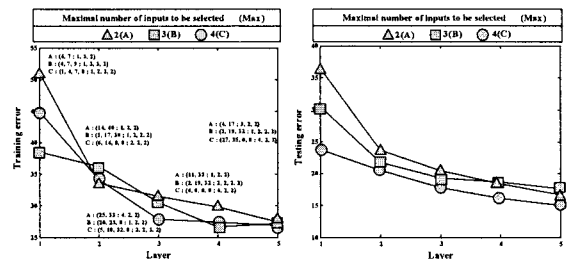
표 2. GAs과 FPNN정보

Parameters	1층~5층	
	Maximum gen	300
Total population size	150	
Selected population size	40	
Crossover rate	0.65	
Mutation rate	0.1	
String length	3+3+30+5	
Maximal no. of inputs to be selected(Max)	$1 \leq l \leq \text{Max}(2-4)$	
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

•; 규칙 후반부 다항식의 입력이 전체 시스템 입력이 사용된 경우

그림 1과 그림 2는 MFs의 형태와 층의 증가에 따른 성능지수를 나타내고 있다. 제시된 그림은 예를 들어 A:(4, 7; 1, 3, 2)가 의미하는 것은 A는 최대로 선택될 수 있는 입력변수의 수를 나타내고, 앞의 두수 4와 7은 선택된 입력변수이다. 그리고 세미콜론 뒤의 숫자 1은 후반부 다항식의 차수를 보여주고 있고, 그 다음의 숫자 3과 2는 4와 7번째 입력변수 각각의 멤버십 함수의 수를 나타내고 있다. 그리고 그림 1은

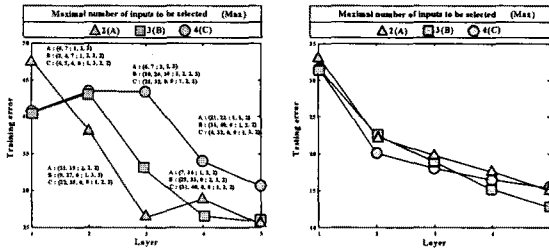


(a) Training error

(b) Testing error

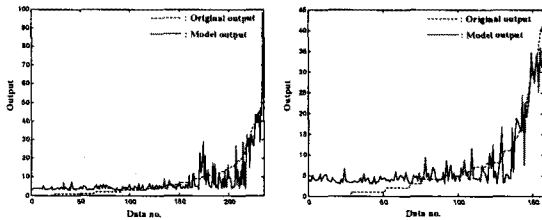
그림 1. Triangular MFs일 때 성능지수

삼각형 멤버십 함수를 그림 2는 가우시안 멤버십 함수일 때 성능지수를 보여주고 있다. 그리고 각 그림의 앞은 training data에 의한 성능지수이고, 뒤의 그림은 testing data에 의한 그림이다.



(a) Training error (b) Testing error
그림 2. Gaussian-like MFs일 때 성능지수

최적의 구조로는 최대 선택될 수 있는 입력변수의 개수 (Max)가 3이며 가우시안형일 때 최적의 모델이 설계되었다. 그림 3은 그때 모델 출력과 실제 출력의 차이를 그래프로 나타내고 있다.



(a) Training output (b) Testing output
그림 3. 실제 출력과 모델 출력의 비교

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=3이며, Gaussian-like 멤버십 함수일 때 최고의 성능 지수를 나타내고 있다. 그림 1과 2에서 보여주듯이 천편일률적인 구조가 아닌 층의 진행에 따라 입력변수 수와 입력변수, 다항식의 차수 멤버십 함수를 최적으로 선택한다.

표 3. 기존의 모델과 제안된 모델의 성능지수비교

Model	Input	M_type	T	Layer	PI	EPI
FPNN [3]	2	T	2	5	32.195	18.462
		G	1	5	49.716	31.423
	3	T	1	5	32.251	19.622
		G	1	5	39.093	19.983
Our model	Max	M_type	T	Layer	PI	EPI
			G	5	25.835	14.375
	2	T	1	5	27.182	17.783
			G	5	25.804	14.637
3	T	1	5	26.235	14.113	
		G	5	26.235	14.113	

T : Triangular, G : Gaussian-like, T : polynomial Type

4. 결론

Advanced GAS-based FPNN 모델은 기존의 FPNN 모델의 최적 구조에 영향을 미치는 FPNN의 입력변수의 수와 후반부 차수, 고정된 입력변수 선택 그리고 멤버십 함수의 수를 설계자의 경험이 아닌 최적 탐색방법인 유전자 알고리즘을 사용하여 입력변수의 수와 그에 따른 입력변수, 후반부 차수 그리고 멤버십 함수를 선택 동조한다. 따라서 설계자의 주관적인 모델이 아닌 객관적인 모델을 만들 수 있고 보다 층에 증가에 따라 유동적인 최적 구조를 설계할 수 있다. 결과적으로 기존의 FPNN의 모델보다 유전자 알고리즘을 사용하여 전반부와 후반부 구조를 선택 동조한 유전론적 최적 퍼지 다항식 뉴럴네트워크의 모델이 성능에 있어서 매우 훌륭한 것을 알 수 있었다.

감사의 글

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

참고 문헌

- [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.
- [7] 오성권, 프로그램에 의한 컴퓨터지능(퍼지, 신경회로망 및 진화 알고리즘을 중심으로), 내하출판사, 2002. 8.
- [8] 오성권, C프로그램에 의한 퍼지모델 및 제어시스템, 내하출판사, 2002.
- [9] 박병준, "진화론적 최적 규칙베이스 퍼지다항식 뉴럴네트워크", 원광대학교 박사학위논문, 2003. 08
- [10] 박호성, "정보 Granule 기반 Fuzzy-Neurocomputing에 의한 고급 컴퓨터 지능 모델에 관한 연구", 원광대학교 박사학위논문, 2005. 02.
- [11] 박건준, "정보 Granules 기반 진화론적 최적 퍼지 추론 시스템의 설계에 관한 연구", 원광대학교 석사학위논문, 2005. 02.
- [12] 진강규, 유전알고리즘과 그 응용, 교우사, 2000