

진화론적 퍼지 다항식 뉴럴 네트워크를 이용한 소프트웨어 공정의 최적 모델 설계

이인태, 오성권, 김현기  
수원대학교 공과대학 전기공학과

Optimal Model Design of Software Process Using Genetically Fuzzy Polynomial Neyral Network

In-Tae Lee, Sung-Kwun Oh, Hyun-Ki Kim  
Department of Electrical Engineering, University of Suwon

**Abstract** - The optimal structure of the conventional Fuzzy Polynomial Neural Networks (FPNN)[3] depends on experience of designer. For the conventional Fuzzy Polynomial Neural Networks, input variable number, number of input variable, number of Membership Functions(MFs) and consequence structures are selected through the experience of a model designer iteratively. In this paper, we propose the new design methodology to find the optimal structure of Fuzzy Polymomial Neural Network by using Genetic Algorithms(GAs)[4, 5]. In the sequel, It is shown that the proposed Advanced Genetic Algorithms based Fuzzy Polynomial Neural Network(Advanced GAs-based FPNN) is more usefual and effective than the existing models for nonlinear process. We used Medical Imaging System(MIS)[6] data to evaluate the performance of the proposed model.

1. 서 론

많은 분야에서 행해지고 있는 시스템 모델링에 관한 연구의 궁극적인 목표는 자연현상을 이해하고, 분석함으로써 원하는 정보를 이용하는데 있다. 그러나 대부분의 현상들은 비선형을 내포하고 있기 때문에 선형성을 바탕으로 개발된 기존의 방법들을 이용해 이들을 해석하기에는 많은 어려움에 부딪히게 된다. 현재까지 개발된 모델들은 입력변수가 증가하면, 또한 시스템이 복잡하고 대규모적이면 함수식으로 표현하는데 한계가 있다.

이러한 문제를 해결하기 위해, A. G. Ivakhnenko는 Group Method of Data Handling(GMDH) [1] 알고리즘을 소개하였다. 그러나 GMDH는 비교적 간단한 시스템에 대해서 매우 복잡한 다항식을 생성하는 경우가 있으며, 2입력 2차 회귀다항식의 형태로 복잡한 비선형 시스템을 표현하는데는 어느정도 한계가 있다. 이를 극복하고자, Oh등에 의해 자기구성 다항식 뉴럴 네트워크(Self-Organizing Polynomial Neural Networks ; SOPNN) [2]와 자기구성 퍼지다항식 뉴럴네트워크(Self-Organizing Fuzzy Neural Networks ; SOFPNN) [3]가 제안되었다. SOFPNN(또는 FPNN)의 구조는 일반적인 다중 퍼셉트론 구조와 달리 미리 정해져 있지 않고 층이 진행되는 동안 모델 스스로 노드의 선택과 제거를 통해 최적의 네트워크 구조를 생성할 수 있다. 그러나 FPNN 구조는 설계자의 경험 또는 반복적인 학습을 통해 노드의 입력변수의 수와 규칙 후반부 다항식 차수, 입력변수 그리고 멤버십의 개수를 선택하였기에 최적의 네트워크 구조를 구축하는데 어려움이 내재되어 있었다.

따라서 본 논문에서는 FPNN을 최적화 시키기 위해 최적화 탐색방법인 유전자 알고리즘(Genetic Algorithms ; GAs)[4, 5]을 이용하여 FPNN의 입력변수 수와 그에

따른 입력변수 그리고 규칙 후반부 다항식의 차수 또한 멤버십 함수의 수를 탐색하여 최적의 FPNN 모델을 구축한다. 평가를 위해 소프트웨어 공정 데이터인 Medical Imaging System (MIS) 데이터[6]를 이용하여 제안된 모델의 정확성 및 예측 성능등을 기존 지능 모델들과 비교를 통해 그 우수성을 보인다.

2. Advanced GAs-based FPNN

2.1 유전자 알고리즘에 의한 구조 동정

본 논문에서 이용한 유전자 알고리즘은 한 개체에 여러 변수의 정보를 담은 직렬 방식을 사용하였고, 선택연산으로는 Roulette-Wheel 방식을, 교차연산으로는 One-crossover방식을 사용하였고, 또한 돌연변이 연산으로는 Invert 방식을 사용하여 최적의 FPNN 구조를 동정하였다. 그 설계과정은 다음과 같다.

[단계 1] GAs-based FPNN 구조를 구축하는데 있어서 기본적인 정보를 결정한다.

[단계 2] GAs-based FPNN 모델을 구축하는데 기본 유닛인 Fuzzy Polynomial Neuron(FPN)[3]을 유전자 알고리즘을 통하여 최적 구조를 선택한다.

[단계 2-1] 입력변수의 수 선택

[단계 2-2] 퍼지규칙 후반부 다항식 차수 선택

[단계 2-3] 입력변수 선택

[단계 2-4] 멤버십 함수의 수 결정

[단계 2-1]에서 [단계 2-4]까지 수행해서 얻은 선택된 입력변수의 수와 퍼지규칙 후반부 다항식 차수, 입력변수의 수에 따른 입력변수 그리고 멤버십 함수의 수를 이용하여 퍼지 추론에 적용한다.

표 1. 퍼지규칙 후반부 다항식의 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

i) 간략 추론

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

$$R^i : \text{If } x_1 \text{ is } A_{\alpha 1} \text{ and } \dots \text{ and } x_n \text{ is } A_{\alpha m} \text{ then } y_i \text{ is } a_{\beta} \quad (1)$$

i번째 데이터의 추론된 값은 다음과 같다.

$$\hat{y}_i = \frac{\sum_{\alpha} w_{\alpha} y_{\alpha}}{\sum_{\alpha} w_{\alpha}} = \frac{\sum_{\alpha} w_{\alpha} a_{\beta}}{\sum_{\alpha} w_{\alpha}} = \sum_{\alpha} \hat{w}_{\alpha} a_{\beta} \quad (2)$$

$$w_{\alpha} = A_{\alpha 1}(x_{1i}) \wedge \dots \wedge A_{\alpha m}(x_{mi})$$

ii) 회귀다항식 추론

후반부가 1, 2, 변형된 2차 다항식 형태로 표현된 것으로 회귀다항식 추론이라 한다. 이 경우의 퍼지 모델은 식(3)과 같은 형태를 가지는 규칙들로 구성된다.

$$R^i: \text{If } x_1 \text{ is } A_{i1} \text{ and } \dots \text{ and } x_n \text{ is } A_{in} \text{ then } y_i = f_i(x_1, \dots, x_n) \quad (3)$$

2.2 유전자 알고리즘을 이용한 FPNN의 설계

[단계 3] 모델의 구축과 테스트 그리고 노드의 선택

Step 1) GAs-based 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(4)와 같이 변환하여 유전자 알고리즘의 적합도 함수로 사용한다.

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

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

Step 5) 구해진 적합도 값( $F_1, F_2, \dots, F_n$ )들을 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 설계하기전 결정된 노드의 제한된 개수 W개 만큼 선호되는 노드를 선택한다.

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

Step 9) 다음 세대에 엘리트 전략을 위해 선택된 노드들 중에서 최적의 적합도 값을 가지는 노드를 선택한다.

Step 10) Step 4)에서 얻어진 정보를 가지고 다음 세대의 population을 생성 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 4] [단계 4]의 현재 층에서 얻어진 최적 노드의 최대 적합도 값  $F_i$ 가 다음 부등식을 만족하는 경우에 알고리즘을 종료한다.

$$F_i \leq F_c \quad (5)$$

단,  $F_c$ 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 5] 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,  $x_{1j}=z_{1i}, x_{2j}=z_{2i}, \dots, x_{wj}=z_{wi}$ 로 구성하고, [단계 2]으로 간다(여기서  $j=i+1$ ). 이 후, [단계 2]부터 [단계 5]까지를 반복한다. 알고리즘이 종료할 경우, 마지막 층의 최적 적합도 값  $F_c$ 를 얻은 노드의 다항식에 입력이 되는 전 층의 출력들을 대입하고 동일한 조작을 제 5층까지 반복하여 최종 추정 모델  $\hat{y}$ 를 얻는다.

3. 실험 데이터를 통한 결과 고찰

3.1 GAs과 FPNN의 기본정보

Advanced GAs-based FPNN의 성능을 테스트하기 위해서 Medical Imaging System (MIS) 데이터[6]를 이용하였다. 이 데이터는 11개의 시스템 입력값과 1개의 출력값으로 구성된 데이터로 본 논문에서 제안한 모델은

11개의 입력변수 중 최대 4가지 선택하도록 설계하였다. 표 2는 각각의 FPNN마다 GAs 정보와 FPNN 모델 구축을 위한 정보를 나타낸 것이다.

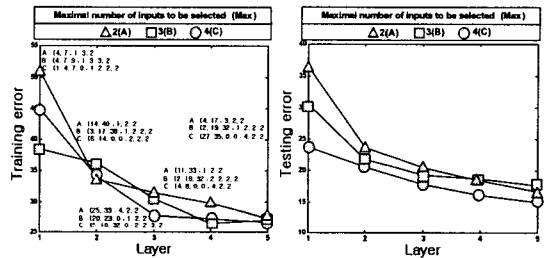
표 2. GAs과 FPNN 정보

Parameters		1층-5층
GA	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
FPNN	Maximal no. of inputs to be selected(Max)	$1 \leq T \leq \text{Max}(2-4)$
	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

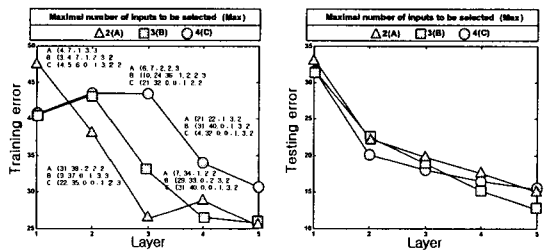
l, T : integer

3.2 Advanced GAs-based FPNN의 성능평가

그림 1과 그림 2는 MFs의 형태와 층의 증가에 따른 성능 지수를 나타내고 있다. 그림에서 A:(4, 7; 1, 3, 2)의 의미는 먼저, A는 Max값을 나타내고 4와 7은 Max 값에 따른 선택된 node 번호를 나타내고 세미콜론(; ) 뒤 첫 번째 숫자 1은 후반부 다항식 형태를 나타내고 그림 3과 2는 각 입력에 대한 MFs의 개수를 나타내고 있다.

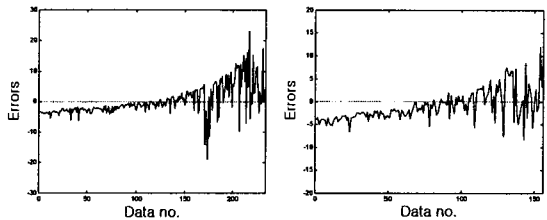


<Training data> <Testing data>  
그림 1. Triangular MFs 일때 성능지수



<Training data> <Testing data>  
그림 2. Gaussian-like MFs 일때 성능지수

그림 1과 2를 살펴보면 층의 증가에 따라 MFs의 숫자와 후반부 타입이 고정되지 않고 유연하게 선택됨을 알 수 있다.



<Training data> <Testing data>  
그림 3. Advanced GAs-based FPNN의 오차곡선

또한 입력변수도 층의 증가에 따라 고정되어 선택되는 게 아니라 가변적으로 선택됨을 볼 수 있다. 그림 3은 실제 출력값과 모델 출력값의 오차곡선이다. 이 그림은 Max=3이고 MFs의 형태가 가우시안일때의 그림으로 데이터 번호에 따른 형태가 다르다.

그림 4는 실제 출력값과 모델에서 구한 출력값을 비교해 주고 있다.

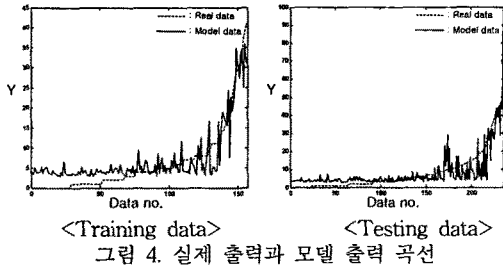


그림 4. 실제 출력과 모델 출력 곡선

그림 5는 Max=3이며 가우시안형일 때, 3층까지의 최적 네트워크 구조를 보여주고 있다. 그림에서 보듯이 각 층별로 입력변수 선택에 있어서 매우 유연하며 각 입력에 대한 멤버십 개수도 유연하게 선택됨을 알 수 있다.

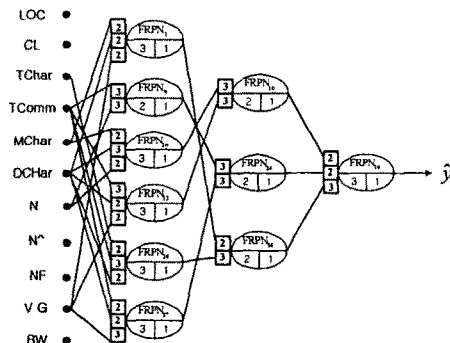


그림 5. Advanced GAs-based FPNN의 최적 구조

기존의 모델과 성능비교를 표 3에서 보여주고 있다. 표에서 보는바와 같이 Max=3이며 MFs의 형태가 가우시안형일 때 PI값과 EPI값이 25.804, 12.637로 성능이 우수함을 볼 수 있다.

표 3. 기존의 모델과 동정오차 비교

Model	Structure			PI	EPI	
SONFN [9]	Simplified	Generic Type	Basic architecture	Case 1	40.753	17.898
	Linear	Generic Type	Basic architecture	Case 2	35.748	17.807
FPNN [3]	Input	M_type	T	2	32.195	18.462
			G	1	49.716	31.423
	3	T	1	32.251	19.622	
		G	1	39.093	19.983	
Our model	2	M_type	T	3	27.406	16.641
			G	1	25.435	14.976
	3	T	1	27.162	17.598	
		G	1	25.804	12.637	
	4	T	4	26.835	14.919	
		G	1	30.69	15.39	

여기에서, Max는 최대도 선택될 수 있는 입력변수를 나타내고 있으며, M\_type은 MFs의 형태를 T는 후반부 다항식 Type을 나타내고 있다.

#### 4. 결 론

본 논문에서 제안한 Advanced GAs-based FPNN 모델은 기존의 FPNN 모델의 최적 구조에 영향을 미치는 네트워크 층, FPNN의 입력변수의 수와 후반부 차수 그리고 고정된 입력변수 선택을 설계자의 경험이 아닌 유전자 알고리즘을 사용하여 입력변수의 수와 그에 따른 입력변수, 후반부 차수 그리고 멤버십 함수의 수를 선택하여 동조한다. 따라서 설계자의 경험이 아닌 객관적이고 유연한 네트워크 구조를 설계할 수 있다. 결과적으로 기존의 모델보다 최적의 네트워크를 찾아내는 시간이 단축되었고 모델의 성능 또한 층의 증가에 따른 유연한 네트워크 구조를 선택 동조하여 향상되었다.

#### 감사의 글

본 연구는 산업자원의 지원에 의하여 기초전력연구원(I-2004-0-074-0-00) 주관으로 수행된 과제임.

#### [참 고 문 헌]

- [1] A. G. Ivakhnenko, "The group method of data handling: a rival of method of stochastic approximation", Soviet Automatic Contr, 1-3, pp. 43-55, 1968.
- [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] J. H. Holland, Adaptation in Natural and Artificial Systems, The University of Michigan Press, Michigan, 1975.
- [5] K. A. De Jong, "Are Genetic Algorithms Function Optimizers?", In R. Manner and B. Manderick, editors, Parallel Problem Solving from Nature 2, North-Holland, Amsterdam, 1992.
- [6] M. R. Lyu, Handbook of Software Reliability Engineering, McGraw-Hill, New York, pp. 510-514, 1995.
- [7] 진강규, 유전알고리즘과 그 응용, 교우사, 2000.
- [8] S. K. Oh and W. Pedrycz and B. J. Park, "Self organizing neurofuzzy networks in modeling software data", Fuzzy set and system, Vol. 145, pp. 165-181, 2004.
- [9] S. K. Oh, W. Pedrycz, and B. J. Park, "Relation-based Neurofuzzy Networks with Evolutionary Data Granulation", Mathematical and Computer Modeling, 2003.
- [10] 박호성, "정보 Granule 기반 Fuzzy-Neurocomputing에 의한 고급 컴퓨터 지능 모델에 관한 연구" 원광대학교 박사학위논문, 2005.
- [11] 박건준, "정보 Granules 기반 진화론적 최적 퍼지 추론 시스템의 설계에 관한 연구", 원광대학교 석사학위논문, 2005.
- [12] 진강규, 유전알고리즘과 그 응용, 교우사, 2000.
- [13] 오성권, C프로그램에 의한 퍼지모델 및 제어시스템, 내하출판사, 2002.
- [14] 오성권, 프로그램에 의한 컴퓨터지능(퍼지, 신경회로망 및 진화 알고리즘을 중심으로), 내하출판사, 2002.