

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

Design of Fuzzy Polynomial neural Networks Using Symbolic Encoding of Genetic Algorithms and Its Application to Software System

이인태¹, 오성권¹, 최정내²

¹ 화성시 봉담읍 수원대학교 전기공학과

E-mail: ohsk@suwon.ackr

² 익산시 신용동 원광대학교 제어계측공학과

E-mail: cin9828@wonkwang.ac.kr

요 약

본 논문은 소프트웨어 공정에 대하여 기호코딩을 이용한 유전자 알고리즘 기반 퍼지 다항식 뉴럴 네트워크 (Genetic Algorithms-based Fuzzy Polynomial Neural Networks ; gFPNN)의 모델을 제안한다. 유전자 알고리즘에는 이진코딩, 기호코딩, 실수코딩이 있다. 제안된 모델은 스트링의 길이에 따른 해밍절벽을 기호코딩으로 극복하였다. gFPNN에 전반부 멤버쉽 함수는 삼각형과 가우시안형의 멤버쉽 함수가 사용된다. 그리고 규칙의 후반부는 간략, 선형, 이차식 그리고 변형된 이차식 함수에 의해 설계된다. 실험적 예제를 통하여 제안된 모델의 성능이 근사화 능력과 일반화 능력이 우수함을 보인다.

Key Words : Symbolic Encoding, Fuzzy Polynomial Neural Networks, Genetic Algorithms, Membership Functions, Optimal Model

1. 서 론

현재 연구자들은 인간의 일상생활 속에서 다양한 지식이나 정보를 사용한 학습, 판단, 예측 등이 인간의 판단과 유사한 결과를 얻을 수 있는 장치를 마련하고자 노력하고 있다. CI 기술은 인간의 신경구조를 모방한 신경망과 인간의 언어적 지식과 전문가적 경험을 사용한 퍼지이론 그리고 유전자 알고리즘과 같은 지능시스템들이 활발하게 연구되면서 합성, 융합, 결합을 통한 고급 CI 기술에 관한 연구가 다양한 형태로 추진되고 있다[1].

이러한 고급 CI 기술 중에 A. G. Inakhnenko가 소개한 Group Method of Data Handling (GMDH) 알고리즘은 비교적 간단한 시스템에 대해서 매우 복잡한 다항식을 생성하는 단점이 있다. 이러한 단점을 극복하고자, Oh 등에 의해 Self-Organizing Polynomial Neural Networks ; SOPNN[2]와 Self-Organizing Fuzzy Polynomial Neural Networks ; SOFPNN[3]가 제안되었다.

SOPNN과 SOFPNN (또는 FPNN)의 구조는 일반적인 다층 퍼셉트론 구조와 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 향상시키기 위하여, 노드의 입력변수의 수와 입력변수, 규칙 후반부 다항식 차수 그리고 멤버쉽 함수의 수를 설계자의 경험 또는 반복적인 학습을 통해 네트워크 구조를 선택하여 최적의 네트워크 구조를 구축하는 데는 어려움이 있었다.

이에 본 논문에서는 고급 CI 기술의 한 영역으로써 효과적인 비선형 문제의 분석과 해결을 위한 Symbolic encoding 기반 유전자적 퍼지 다항식 뉴럴네트워크(Symbolic encoding based Genetically Fuzzy Polynomial Neural Networks; gFPNN)의 설계를 목표로 한다. FPNN을 최적화시키기 위해서 최적화 탐색 방법인 기호 코딩 기반 유전자 알고리즘[4, 5]을 이용하여 FPNN의 입력변수 수와 입력변수, 규칙 후반부 다항식의 차수 그리고 멤버쉽 함수의 수를 탐색하여 최적의 퍼지 다항식 뉴럴

네트워크를 구축한다. 이전 이진코딩에서 발생 하는 두 수 사이의 급격한 비트 차이에 의한 해밍절벽을 해결하기 위하여 기호코딩을 사용 하였다. 평가를 위해서 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. 퍼지 다항식 뉴런 (FPN)

FPN은 FPNN의 가장 기본이 되는 처리 단 위의 일반적인 형태로써, 퍼지 집합과 뉴럴 네트워크의 개념에 기초해서 자세히 설명된다. 그림 1에서 보여진 것처럼 FPN은 2개의 기본적인 기능 모듈로 구성되어 있다(여기서 그림 1은 예를 들어 2입력이고 멤버십 함수의 수는 2x2인 경우이다). 처음 하나는 "F" 로 표기되어진 것으로, 이는 퍼지 집합의 집합체로써 입력 수의 변수들과 뉴런에 의해 실현된 처리 부분사이를 상호 연결해 준다. 그림에서 X_p 와 X_q 를 입력변수들로 표시한다. "P" 로 표기된 두 번째 모듈은 함수 기반 비선형(다항식) 처리에 관한 것이다.

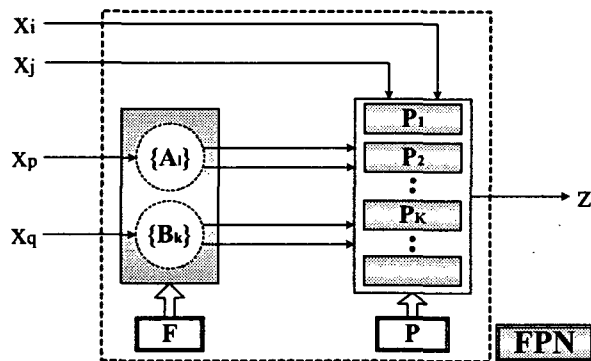


그림 1. 퍼지 다항식 뉴런 (FPN)의 구조

3. 기호코딩 기반 gFPNN의 설계

기호코딩 기반 유전자적 퍼지 다항식 뉴럴네트워크의 최적화 설계과정을 다룬다.

[단계 1] 출력변수 y 에 관계하는 n 개의 시스템 입력변수를 결정한다.

[단계 2] N 개의 입출력 데이터를 학습용 데이터와 평가용 데이터로 분할한다.

[단계 3] gFPNN 구조를 구축하는데 있어서 기본적인 정보를 결정한다. 즉, gFPNN의 퍼지 추론 방법 및 동정을 설계하는 부분과 모델의 구조를 결정하는 초기 정보를 표2와 같이 결정한다.

다.

[단계 4] gFPNN 모델을 구축하는데 기본 유닛인 FPN 노드들을 유전자 알고리즘에서 각각의 염색체에 의해 구조를 선택한다.

[단계 4-1] 입력변수의 수 선택

[단계 4-2] 입력변수 선택

[단계 4-3] 멤버십 함수의 수 선택

[단계 4-4] 퍼지규칙 후반부 다항식 차수 선택

[단계 4-1]부터 [단계 4-4]까지 수행해서 얻은 선택된 입력변수와 퍼지규칙 후반부 다항식 차수 그리고 멤버십 함수의 수를 이용하여 퍼지 추론을 적용한다.

gFPNN의 구조 동정을 위한 염색체의 구조는 다음과 같다.

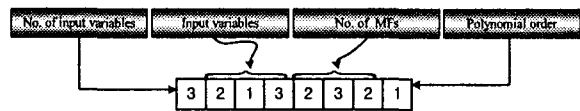


그림 2. Symbolic encoding에 의한 염색체 구조

표 1. 퍼지규칙 후반부에서 회귀다항식의 서로 다른 형태

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

[단계 5] 모델의 구축과 테스트 그리고 노드의 선택

Step 1) gFPNN 모델 구축을 위한 유전자의 세대, Population 개수, 돌연변이율, 교배율, 그리고 염색체의 길이와 같은 유전자의 정보를 설정한다.(표 2)

Step 2) 첫 번째 세대에서 설정된 population 개수만큼 염색체에 의한 노드(FPN)를 구축한다.

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

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

Step 4) 다음 세대의 생성을 위해서 구해진 개체 구 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 설계하기전 결정한 노드의 제한된 개수 W개만큼 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 가진 노드를 선택한다.

Step 10) Step 4)에서 얻어진 정보를 가지고 다음 세대의 population을 생성한 후 Step 2)에서 Step 9)까지 반복한다. 여기서 Step 3)에서 생성된 적합도 값들 중에서 가장 낮은 적합도 값을 가진 노드를 Step 9)에서 구한 이전 세대의 최적의 적합도 값을 가진 노드로 대체하는 엘리트 전략을 실행한다.

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 6] [단계 5]의 현재 층에서 얻어진 최적 노드의 최대 적합도 값 F1이 다음 부등식을 만족하는 경우에 알고리즘을 종료한다.

$$F1 \leq F^* \quad (2)$$

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

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

[단계 7] 현재 층에서 보존된 노드의 출력들(z1i, z2i, ..., zWi)에서 다음 층의 새로운 입력들(x1j, x2j, ..., xWj)로, x1j=z1i, x2j=z2i, ..., xWj=zWi 으로 구성하고, [단계 4]로 간다(여기서, j=i+1). 이 후, [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 1층 까지 반복하여 최종 추정 모델 \hat{y} 를 얻는다.

4. 실험 데이터를 통한 결과 고찰

제안된 모델의 성능을 테스트하기 위해서 Medical Imaging System (MIS) 데이터[6]를 이용하였다. 이 공정은 기존의 이용할 수 있는 데이터들 보다 풍부한 데이터들로 구성되어 있

기 때문에 모델을 평가하는데 많은 도움이 된다. 표 2는 GAs정보와 FPNN 모델구축을 위한 정보를 나타낸 것이다.

표 2. GAs와 FPNN정보

	Parameters	1층~3층
GAs	Maximum gen	150
	Total population size	300
	Selected population size	40
	Crossover rate	0.65
	Mutation rate	0.1
	String length	1+(Max*2)+1
FPNN	Maximal no. of inputs to be selected(Max)	1 ≤ l ≤ Max(2~5)
	polynomial Type(Type T) of the consequent part of rules	1 ≤ T ≤ 4
	Membership Function(MFs) type	Gaussian Triangular
	No. of MFs per each input	2 or 3

l, T ; interger

* ; 규칙 후반부 다항식의 입력이 전체 시스템 입력이 사용된 경우

Max ; 최대로 선택될 수 있는 입력변수의 수

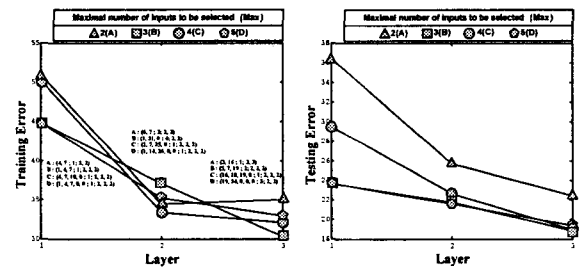


그림 3. 층별 gFPNN의 성능지수(삼각형 MFs)

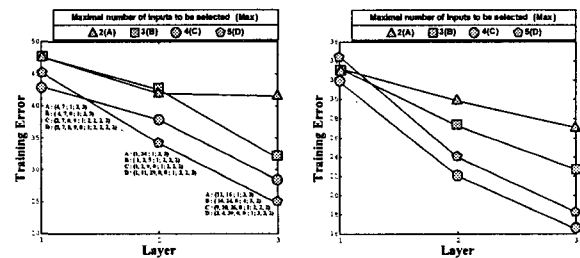


그림 4. 층별 gFPNN의 성능지수(가우시안형 MFs)

그림 3과 4는 후반부 다항식에 선택입력을 사용했을 때의 최대 입력(Max)값에 따른 각 층별 성능지수를 나타내고 있다. 그림에서 A(4, 7 ; 1; 3, 3)을 풀이하면 최대 입력이 2일 때(A) 선택된 입력변수는 4번과 7번입력변수이고 후반부 다항식은 간략(1)을 선택하고 선택된 입력변수에 해당하는 멤버쉽 함수의 개수는 각각 3개 선택을 보여주고 있다. 그림 3과 4는 앞의 그림이 PI(training data)값에 의한

그림이고 뒤의 그림이 EPI(testing data)에 의한 그림이다.

최적화된 GAs 기반의 FPNN 모델 구조를 구축할 수가 있었다.

표 3 기존의 모델과 동정오차비교

	Max	M_type	T	PI	EPI
FPNN [3]	2	T	2	34.169	25.997
		G	1	47.404	32.535
	3	T	1	34.637	21.753
		G	1	37.665	21.364
Our model	2	T	1	35.026	22.402
		G	1	40.444	27.090
	3	T	2	30.326	18.854
		G	4	31.808	22.678
	4	T	1	32.052	18.976
		G	1	28.417	16.522
	5	T	3	32.993	19.350
		G	1	24.724	18.270

T : Triangular , G : Gaussian-like

표 3은 MIS 데이터[6]를 사용하여 성능지수를 식 (3)에 의해 계산된 값으로 기존의 모델과의 성능비교를 한 것이다. Max=4 이고 멤버쉽 함수가 가우시안형이며 Type 1일 때 PI=28.417과 EPI=16.522로 성능이 우수함을 보인다. 또한 기존의 FPNN보다 gFPNN이 Max 값에 따른 입력변수의 선택과 후반부 다항식의 선택 그리고 멤버쉽 함수의 수 선택에 있어서 일률적이지 않고 유동적으로 최적의 모델을 찾는데 우수하였다.

5. 결론

본 논문에서는 기호코딩 기반 gFPNN의 효율적인 설계방법을 제안하였다. 기존의 FPNN 구조는 성능을 향상시키는데 크게 영향을 미치는 네트워크 층, 노드내의 입력변수의 수와 퍼지규칙 후반부 다항식의 차수 그리고 전반부 멤버쉽 함수의 수 선택에 있어서 설계자의 경험에 의존하여 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, 후반부 다항식의 차수 그리고 멤버쉽 함수의 수를 선택 동조하였다. 그리고 이진코딩에서 문제가 되었던 스트링의 길이가 길면 길수록 인접한 두 수 사이에 발생하는 급격한 비트 차이에 의한 해밍절벽을 극복하기 위해 기호코딩을 사용하여 보다 최적화된 구조 동정을 할 수 있게 하였다. 따라서 제안된 모델은 모델 구축에 있어서 유연성과 정확성을 가지며 더욱 정확한 예측 능력을 가진

감사의 글

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

참고 문헌

- [1] W. Pedrycz and J.F. Peters, Computational Intelligence and Software Engineering, World Scientific, Singapore, 1998.
- [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 Arbor. 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] S. K. Oh, W. Pedrycz, and B. J. Park, "Relation-based Neurofuzzy Networks with Evolutionary Data Granulation", Mathematical and Computer Modeling, 2003.
- [8] S.-K. Oh and W. Pedrycz, "Self-organizing Polynomial Neural Networks Based on PNs or FPNs : Analysis and Design", Fuzzy Sets and Systems, volume 142, pp163-198, 1 March 2004.