정보 및 제어 심포지엄(ICS '05) 2005. 5. 14

유전자 알고리즘 기반 퍼지 다항식 뉴럴네트워크를 이용한 비선형 공정데이터의 최적 동정

Optimal Identification of Nonlinear Process Data Using GAs-based Fuzzy Polynomial Neural Networks

이인태*, 김완수**, 김현기***, 오성권****
(In-Tae Lee, Wan-Su Kim, Hyun-Ki Kim, Sung-Kwun Oh)

Astract - In this paper, we discuss model identification of nonlinear data using GAs-based Fuzzy Polynomial Nerual Networks(GAs-FPNN). Fuzzy Polynomial Neural Networks(FPNN) is proposed model based Group Method Data Handling(GMDH) and Neural Networks(NNs). Each node of FPNN is expressed Fuzzy Polynomial Neuron(FPN). Network structure of nonlinear data is created using Genetic Algorithms(GAs) of optimal search method. Accordingly, GAs-FPNN have more inflexible than the existing models (in)from structure selecting. The proposed model select and identify its for optimal seach of Genertic Algorithms that are no. of input variables, input variable numbers and consequence structures. The GAs-FPNN model is select tuning to input variable number, number of input variable and the last part structuer through optimal seach of Genertic Algorithms. It is shown that nonlinear data model design using Genetic Algorithms based FPNN is more usefulness and effectiveness than the existing models.

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

1. 서론

대부분의 자연현상들은 근본적으로 비선형을 내포하고 있기 때문에 선형성을 바탕으로 개발된 기존의 방법들을 이용해 이들을 해석하기에는 많은 어려움에 부딪히게 된다. 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, 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 표현하는데는 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 진행하는 동안 모델 스스로 노드의 선택과 제거를 통해 최적의 네트워크 구조를 생성할 수 있다. 그러나 기존 FPNN 구조는 성능을 향상시키기 위하여, 노드의 입력변수의 수와 규칙 후반부 다항식 차수 그리고 입력변수는 설계자의 경험 또는 반복적인 학습을 통해 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 입력변수 수와 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. GAs-based FPNN 설계

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

[단계 1] 출력변수 y에 관계하는 n개의 시스템 입력변수를 결정한다. N개의 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

* 準 會 員 : 水原大學校 電氣工學科 碩士課程
** 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 敎授
**** 正 會 員 : 水原大學校 電氣工學科 敎授

[단계 3-2] 퍼지규칙 후반부 다항식 차수 선택
[단계 3-3] 입력변수 선택
[단계 3-1]부터 [단계 3-3]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, 돌연변이률, 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)와 같이 변환하여 유전자 알고리즘의 적합도 함수로 사용한다.

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

Step 4) 다음 세대의 생성을 위해서 구해진 개체구 정보와 적합도 값들을 이용하여 선택연산, 교배, 돌연변이를 실행한다.
Step 5) 구해진 적합도 값들(F_1, F_2, …, 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 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_1이 다음 부등식을 만족하는 경우에 알고리즘을 종료한다.

$$F_1 \le F_*. \quad (2)$$

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

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

[단계 6] 현재 층에서 보존된 노드의 출력들(z_{1i}, z_{2i}, …, z_{Wi})에서 다음 층의 새로운 입력들(x_{1j}, x_{2j}, …, x_{Wj})로, $x_{1j}=z_{1i}$, $x_{2j}=z_{2i}$, …, $x_{Wj}=z_{Wi}$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 제 5층까지 반복하여 최종 추정 모델 $\hat{y}$를 얻는다.

3. 실험 데이터를 통한 결과 고찰

GAs-based FPNN의 성능을 테스트하기 위해서 MIS데이터[6]를 이용하였다. 이 공정은 기존의 이용할 수 있는 데이터들 보다 풍부한 데이터들로 구성되어 있기에 모델을 평가하는데 많은 도움이 된다. 표 2는 각각의 FPN마다 GA 정보와 퍼지 다항식 뉴럴 네트워크 구축을 위한 정보를 나타낸 것이다.

표 2. GAs과 FPNN정보

	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)	$1 \le l \le$ Max(2~5)
	polynomial Type(Type T) of the consequent part of rules	$1 \le T_* \le 4$
	Membership Function(MFs) type	Gaussian / Triangular
	No. of MFs per each input	2

l, T ; interger
$*$; 규칙 후반부 다항식의 입력이 전체 시스템 입력이 사용된 경우

그림 1과 그림 2는 MFs의 형태와 층의 증가에 따른 성능지수를 나타내고 있다. A:(3 ; 3, 11)가 의미하는 것은 A는 Max값, 앞의 3은 후반부 차수이며 뒤에 3과 11은 선택되어진 최적의 node번호이다.

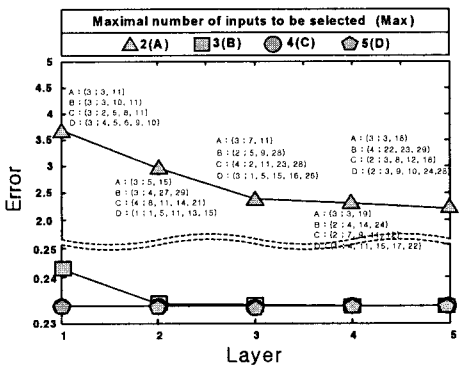

그림 1 Triangular MFs일때 성능지수

그림 1과 2에서 Max값이 3이상이며 층마다 다른 후반부 차수를 선택해야 최적의 모델 설계가 용이하다는 것을 보여

주고 있다. 그리고 표3을 살펴보면 제안된 모델이 기존의 모델보다 성능도 뛰어나고 모델의 최적화 과정도 보다 유연함을 볼 수 있다.

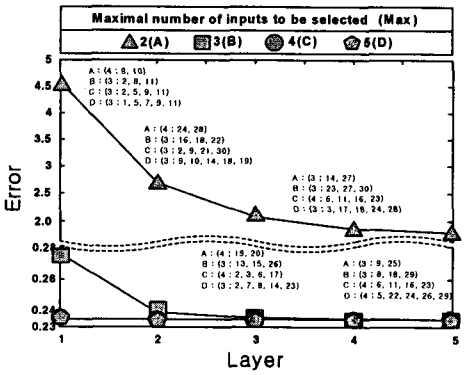

그림 2 Gaussian-like MFs일때 성능지수

최적의 구조로는 입력변수의 개수가 5입력이며 가우시안형일 때 그리고 후반부 차수가 선형 또는 변형된 2차식일 때 최적의 모델을 설계하였다. 그림 3은 모델 출력값과 실제 출력값의 차이를 그래프로 나타내고 있다.

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, T : polynomial Type

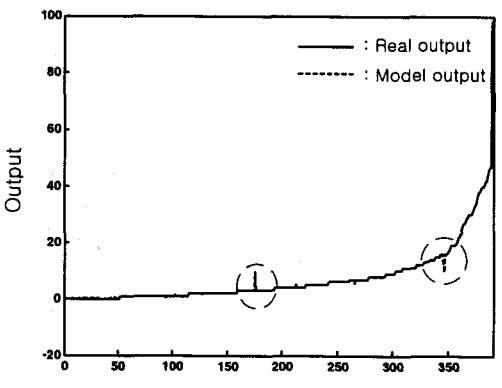

그림 3 모델의 출력과 실제 출력의 오차

4. 결론

GAs-based FPNN 모델은 기존의 FPNN 모델의 최적 구조에 영향을 미치는 네트워크 층, FPN의 입력변수의 수와 후반부 차수 그리고 고정된 입력변수 선택을 설계자의 경험이 아닌 최적 탐색방법인 유전자 알고리즘을 사용하여 입력변수의 수와 그에 따른 입력변수 그리고 후반부 차수를 선택 동조한다. 따라서 설계자의 주관적인 모델이 아닌 객관적인 모델을 만들 수 있고 보다 유연한 최적 구조를 설계할 수 있다. 결과적으로 기존의 FPNN의 모델보다 유전자 알고리즘을 사용하여 전반부와 후반부 파라미터를 선택 동조하는 것이 최적화된 모델 설계에 유리한 것을 알 수 있다. 하지만 최적 모델 설계에 영향을 미치는 MFs 형태와 개수를 선택하는 과제가 남아있다. 이는 차후에 최적 탐색 알고리즘을 확장시켜 모델 설계 방법을 연구 할 예정이다.

감사의 글

본 연구는 산업자원부의 지원에 의하여 기초전력연구원 (R-2003-0-285) 주관으로 수행된 과제임.

참 고 문 헌

[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, "Handbood 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