

# 비선형 공정을 위한 최적 다항식 뉴럴네트워크에 관한 연구

## A Study on Optimal Polynomial Neural Network for Nonlinear Process

김 완 수\* · 오 성 권\*\* · 김 현 기\*\*\*

(Wan-Su Kim · Sung-Kwun Oh · Hyun-Ki Kim)

**Abstract** - In this paper, we propose the Optimal Polynomial Neural Networks(PNN) for nonlinear process. The PNN is based on Group Method of Data Handling(GMDH) method and its structure is similar to feedforward Neural Networks. But the structure of PNN is not fixed like in conventional Neural Networks and can be generated. The each node of PNN structure uses several types of high-order polynomial such as linear, quadratic and modified quadratic, and is connected as various kinds of multi-variable inputs. The conventional PNN depends on experience of a designer that select No. of input variable, input variable and polynomial type. Therefore it is very difficult a organizing of optimized network. The proposed algorithm identified and selected No. of input variable, input variable and polynomial type by using Genetic Algorithms(GAs). In the sequel the proposed model shows not only superior results to the existing models, but also pliability in organizing of optimal network. Medical Imaging System(MIS) data is simulated in order to confirm the efficiency and feasibility of the proposed approach in this paper.

**Key Words** : Polynomial Neural Networks(PNN), Group Method of Data Handling(GMDH), Genetic Algorithms(GAs), Medical Imaging System(MIS) data, Mean Squared Error(MSE)

### 1. 서 론

인간의 문명이 발달함에 따라 시스템이 복잡해지고 대규모 구조로 변하여 시스템을 모델링 하는데 많은 어려움을 겪고 있다. 시스템이 복잡하고 대규모 구조인 경우에는 함수식으로 모델의 표현이 한정되지 않기 때문에 회귀분석을 적용할 수 없는 경우도 있다. 즉, 비선형 시스템의 동적 분석을 표현하는데 있어서의 수학적 모델들은 많은 입·출력 변수들 사이에서 모델을 구축하는 변수들과 모델 구조의 선택에 많은 문제를 가지고 있다. 이러한 문제를 해결하기 위한 한 가지 방법으로 A. G. Ivakhnenko는 Group Method of Data Handling (GMDH) 알고리즘을 제안하였다[1-3]. GMDH는 2변수 2차식에 의한 부분표현식을 계층적으로 조합하여 비선형 모델 추정식을 얻을 수 있다. 그러나 자연계의 다양한 시스템을 예측하는데 있어 2변수로 한정을 하여 다양한 입력변수들끼리의 상호의존적인 면을 약하게 하였으며, 2차식으로는 다소 복잡한 시스템을 표현하기에는 역부족이다.

이를 개선하기 위해 S-K. OH는 Self-Organizing Polynomial Neural Networks(SOPNN 또는 PNN)을 설계하였다[4, 5]. SOPNN 모델은 입력변수의 선택과 입출력 데이

터의 분할, 부분표현식을 정의하여 시스템 방정식을 예측한다. 즉 입력변수를 2변수로 한정짓지 않고 설계자가 원하는 수만큼 선택할 수 있게 설계하였다. 하지만 이 모델 또한 선화된 모델을 구하기 위해서는 노드의 입력변수의 수, 입력 변수 그리고 다항식 차수를 설계자의 시행착오에 의존하여 결정해야만 한다. 따라서 최적의 성능을 얻기 위하여 각 노드의 부분표현식을 구성하는 입력변수의 수, 입력변수 그리고 다항식 차수를 결정하는 객관적이고 신뢰성 높은 새로운 최적화 방법이 필요하다.

이에 본 논문에서는 최적탐색 방법인 유전자 알고리즘[6, 7]을 이용하여 입력변수의 수와 그에 따른 입력변수 그리고 다항식의 차수를 탐색하는 Genetic Algorithms-based Polynomial Neural Networks(GAs-based PNN)를 제안한다. 이 모델은 설계자의 경험에 의존하지 않고 유전자 알고리즘을 이용하여 입력변수의 수, 입력변수 그리고 다항식 차수를 선택한다.

본 논문은 GAs-based PNN 구조로 최적의 모델을 구축하고, 비선형 공정에 적용하기 위하여 Medical Imaging System(MIS)[8] data를 사용하여 제안된 모델의 타당성, 정확성 및 예측 성능 등을 보인다.

### 2. 진화론적 최적 다항식 뉴럴 네트워크 모델

유전자 알고리즘을 이용한 다항식 뉴럴 네트워크의 최적화 설계방법은 다음과 같이 단계별로 설명된다.

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

\* 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 · 工博  
\*\*\* 正 會 員 : 水原大學校 電氣工學科 教授 · 工博  
接受日字 : 2005년 10월 10일  
最終完了 : 2005년 10월 28일

[단계 2] 전체 입출력 데이터를 학습용 데이터와 테스트용 데이터로 분할한다.

[단계 3] PNN 구조를 구축하기 위한 기본적인 정보를 결정한다.

[단계 3-1] 종료조건을 선택 - 최적 노드의 적합도값 비교 또는 최대 층 수 결정

[단계 3-2] 노드에 입력될 최대 입력변수의 수 결정

[단계 3-3] 하나의 층에서 생성될 노드의 수 결정

[단계 3-4] 목적함수의 하중계수 값을 결정

[단계 4] 유전자 알고리즘을 이용한 PNN 구조 결정 및 각 노드의 입력변수 수와 회귀 다항식 차수를 결정한다.

[단계 4-1] 입력변수의 수 선택

[단계 4-2] 다항식 차수 선택

[단계 4-3] 입력변수 선택

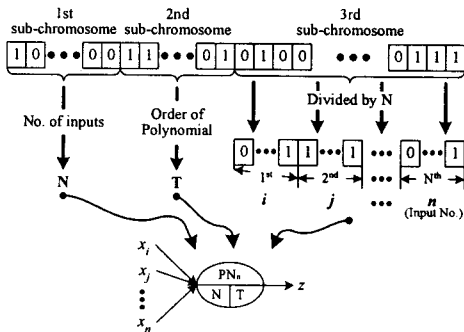


그림 1 유전자 알고리즘의 염색체를 이용한 PN 설계

표 1 회귀다항식의 차수 형태

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

[단계 5] 유전자 알고리즘에 의한 최적 모델의 구축과 테스트 그리고 노드를 선택한다. 선형 회귀분석에는 학습 데이터를 사용하고, 성능지수 즉 동정오차는 다음의 식 (1)을 이용한다.

$$E = \frac{1}{N_t} \sum_{m=1}^{N_t} \{y_m - \hat{y}_m\}^2 \quad (1)$$

여기서,  $N_t$  는 학습데이터 총 수이다.

추정한 후반부 다항식에 테스트 데이터를 사용하여 식(1)의 테스트 데이터에 대한 동정오차를 계산한다.

[단계 6] 단계 5의 현재 층에서 얻어진 최소 동정오차  $E_1$  이 다음의 부등식을 만족하는 경우에 알고리즘을 종료한다.

$$E_1 \geq E_n \quad (2)$$

단,  $E_n$  는 이전 층의 최소 동정오차 이다.

[단계 7] 현재 층에서 보존된 노드의 출력들에서 다음 층의 새로운 입력들이 구성되고, [단계 4]부터 [단계 7]까지 반복한다.

본 논문에서는 복잡한 구조를 피하기 위해 식(2)의 종료 판정을 사용하지 않고, 5층으로 제한하였다.

### 2.1 하중계수를 가진 목적 함수

본 논문에서는 근사화 및 일반화의 상호 균형과 의존능력을 가진 합성 목적함수(성능지수)를 이용하여 상호 연계된 최적 모델 구조를 생성한다[9].

$$f(PI, EPI) = \theta \times PI + (1 - \theta) \times EPI \quad (3)$$

본 논문에서는 식(3)의 하중계수 값이  $\theta=0.5$  인 경우를 고찰 하도록 한다.

### 3. 시뮬레이션 및 결과 고찰

비선형 공정 및 다변수 입력 시스템에 적용하기 위하여 Medical Imaging System(MIS)[8] data를 사용하였다. MIS data는 총 390개의 소프트웨어 모듈들의 부분집합으로 11입력( $LOC, CL, TChar, TComm, MChar, DChar, N, \hat{N}, N_F, V(G), BW$ )과 1출력( $Changes$ )을 가지는 data이다. 따라서, MIS data는  $390 \times 12$  매트릭스 형태가 된다. MIS data의 경우, 성능지수는 식(1)의 Mean Squared Error(MSE)를 이용한다.

전체  $390 \times 12$ 개 데이터 중에서 검증된 기준[8]에 의해서 다음과 같이 Training Data와 Testing Data를 분류하고 각각에 대하여 입력과 출력으로 분류한다.

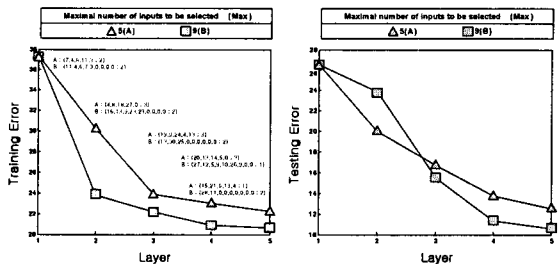
#### [Training Data]

총 390개의 소프트웨어 모듈들 중에 234개의 모듈을 선택하여 입력  $234 \times 11$ , 출력  $234 \times 1$  로 분류한다.

#### [Testing Data]

총 390개의 소프트웨어 모듈들 중에 156개의 모듈을 선택하여 입력  $156 \times 11$ , 출력  $156 \times 1$  로 분류한다.

그림 2는 하중 계수 값  $\theta=0.5$ 일 때, 출력  $\hat{y}$ 의 성능지수를 나타낸 것이다. 그림에서  $\Delta 5(A)$ 는 5입력이고,  $\square 9(B)$ 는 9입력을 나타낸다. 그리고 각각의 입력에 대하여 A(최적노드; 다항식 차수), B(최적노드; 다항식 차수)를 나타낸다.



(a) Training error

(b) Testing error

그림 2 출력  $\hat{y}$ 의 성능지수 ( $\theta=0.5$ )

그림 2에서 알 수 있듯이 층이 증가할수록 모델의 근사화 와 일반화 능력이 뚜렷하게 향상되는 것을 알 수 있다.

그림 3은 하중계수 값이  $\theta=0.5$ 이고 최대입력이 9인 경우 에 3층에서 네트워크를 나타낸 것이다.

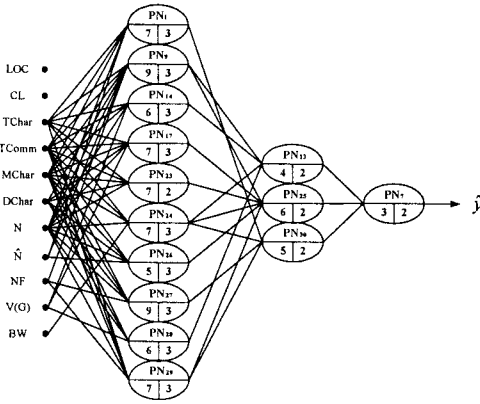


그림 3 GAs-based PNN의 최적 구조 ( $\theta=0.5$ )

그림 4는 최적 구조일 때 모델 출력과 실제 출력의 차이 를 그래프로 나타내고 있다.

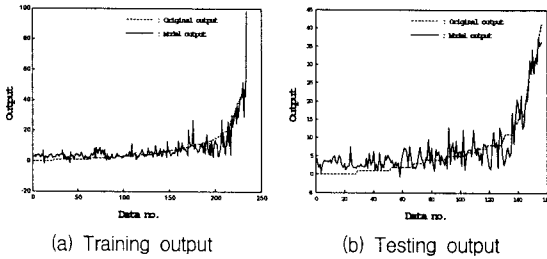


그림 4 실제 출력과 모델 출력의 비교

표 2는 MIS data를 사용하여 얻어진 기존 PNN 모델과 본 논문에서 제안된 GAs-based PNN 모델의 동정오차비교 이고, 표 3은 하중계수 값이  $\theta=0.5$ 일 때, training data와 testing data를 사용하여 얻어진 동정오차비교이다.

표 2 MIS data를 사용한 동정오차비교

Model	Input Max	Layer	Type	Performance Index
PNN	9	5	3	8.456
Our Model	9	5	2	7.843

표 3 Training data와 Testing data를 사용한 동정오차비교

Model	Input Max	Layer	Type	Performance Index	Performance Index
PNN	5	5	4	25.709	15.730
Our Model	5	5	1	22.238	12.566
	9	5	2	20.654	10.621

#### 4. 결 론

본 논문에서는 유전자 알고리즘을 이용하여 다항식 뉴럴 네트워크의 최적 구조를 설계하였다. 기존의 PNN 알고리즘 은 층을 구성하는 노드의 입력변수의 수와 다항식 차수 그리고 고정된 입력변수 선택을 설계자의 경험에 의해서 결정 함으로써 객관적이고 주어진 데이터의 특성에 맞는 최적의 모델을 구축하지 못하였고, 또한 최적의 모델을 구축하기 위 해서 많은 시행착오를 필요로 하였다. 이러한 기존의 PNN 모델의 구조의 한계를 극복하기 위하여 최적화 방법인 유전 자 알고리즘을 이용하여 노드의 입력변수의 수와 그에 해당 한 입력변수 그리고 다항식 차수 등의 파라미터를 최적 선택하여 모델 구축에 있어서 유연성과 정확성을 가지며 더욱 정확한 예측 능력을 가진 최적화된 PNN 모델구조를 구축하 였다. 이러한 최적화된 GAs-based PNN 구조를 비선형 공 정에 적용하기 위하여 비선형 다변수의 MIS data를 통하여 성능을 확인하였다.

#### 감사의 글

본 연구는 산업자원부의 지원에 의하여 기초전력 연구원 (R-2003-B-274) 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 Control, 1-3, pp. 43-55, 1968
- [2] A. G. Ivakhnenko, "Polynomial theory of complex systems", IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-1, pp. 364-378, 1971
- [3] A. G. Ivakhnenko and H. R. Madala, Inductive Learning Algorithms for Complex Systems Modeling, CRC Press, London, 1994
- [4] S. K. Oh and W. Pedrycz, "The design of self-organizing Polynomial Neural Networks", Information Science, Vol. 141, pp. 237-258, 2002
- [5] S. K. Oh and W. Pedrycz and B. J. Park, "Polynomial Neural Networks Architecture : Analysis and Design", Computers and Electrical Engineering, 2002(in press)
- [6] Holland, J. H., Adaptation In Natural and Artificial Systems. The University of Michigan Press, Ann Arbour. 1975
- [7] D. E. Goldberg, Genetic Algorithm in search, Optimization & Machine Learning, Addison wesley, 1989
- [8] MICHAEL, R. L.: 'Handbook of Software Reliability Engineering' McGraw-Hill, pp. 510-514, 1995
- [9] S. K. Oh and W. Pedrycz, "Identification of Fuzzy Systems by means of an Auto-Tuning Algorithm and Its Application to Nonlinear Systems", Fuzzy sets and Systems, Vol. 115, No. z, pp. 205-230, 2000