

유전자 알고리즘 기반 최적 다항식 뉴럴네트워크 및 비선형 공정으로의 응용

Genetic Algorithms based Optimal Polynomial Neural Network and Its application to Nonlinear Process

김완수, 오성권, 김현기
수원대학교 전기공학과

Wan-Su Kim, Sung-Kwun Oh, Hyun-Ki Kim
Dept. of Electrical Engineering, The University of Suwon
E-mail : ohsk@suwon.ac.kr

요 약

본 논문은 최적 탐색 알고리즘인 유전자 알고리즘을 이용하여 다항식 뉴럴네트워크 (Polynomial Neural Networks : PNN)의 최적 설계가 그 목적이다. 기존의 다항식 뉴럴네트워크는 확장된 GMDH(Group Method of Data Handling) 방법에 기반을 두며, 네트워크의 성장과정을 통하여 각 층의 다항식뉴런(혹은 노드)에서 고정된 (설계자에 의해 미리 선택된) 노드 입력들의 수뿐만 아니라 다항식 차수(1차, 2차, 그리고 수정된 2차식)를 이용하였다. 더구나, 그 방법은 학습을 통해 생성된 PNN이 최적 네트워크 구조를 가진다는 것을 보증하지 못한다. 그러나, 제안된 GA-based PNN 모델은 다음의 파라미터들- 즉 입력변수의 수, 입력변수, 및 다항식 차수-을 유전자 알고리즘을 이용하여 선택 동조함으로써 그 구조를 구조적으로 더 최적화된 네트워크가 되도록 하고, 기존의 PNN보다 훨씬 더 유연하고, 선호된 뉴럴 네트워크가 되도록 한다. 하중계수를 가진 합성성능지수가 그 모델의 근사화 및 일반화(예측) 능력 사이의 상호 균형을 얻기 위해 제안된다. GA-based PNN의 성능을 평가하기 위해 그 모델은 가스 터빈 발전소의 NOx 배출 공정 데이터로 실험된다. 비교해석은 제안된 GA-based PNN이 앞서 나타난 다른 지능모델보다 더 우수한 예측능력뿐만 아니라 높은 정확성을 가진 모델임을 보인다.

1. 서론

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 데이터의 분할, 부분표현식을 정의하여 시스템 방정식을 예측한다. 즉 입력변수를 2변수로 한정짓지 않고 설계자가 원하는 수만큼 선택할 수 있게 설계하였다. 하지만

이 모델 또한 선호된 모델을 구하기 위해서는 노드의 입력변수의 수, 입력변수 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 Algorithm-based Polynomial Neural Networks(GA-based PNN)를 제안한다. 이 모델은 설계자의 경험에 의존하지 않고 유전자 알고리즘을 이용하여 입력변수의 수, 입력변수 그리고 다항식 차수를 선택한다. 본 논문은 GA-based PNN 구조로 최적의 모델을 구축하고, 비선형 공정에 적용하기 위하여 가스터빈 발전소의 NOx 방출 공정 데이터[8]를 사용하여 제안된 모델의 타당성, 정확성 및 예측 성능 등을 보인다.

2. GA-based 다항식 뉴럴 네트워크

2.1 GA-based 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.
- [단계 2] 전체 입출력 데이터를 학습용 데이터와 테스트용 데이터로 분할한다.
- [단계 3] PNN 구조를 구축하기 위한 기본적인 정보를 결정한다.
 - [단계 3-1] 종료조건 선택 - 최적 노드의 적합도 값 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.

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

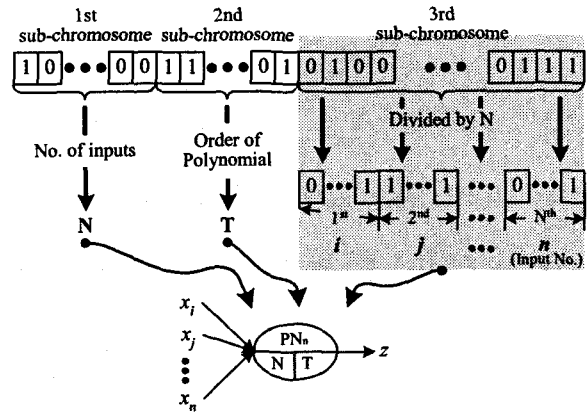


그림 1 GA의 염색체를 이용한 PN 설계

- [단계 4-1] 입력변수의 수 선택
- [단계 4-2] 다항식 차수 선택
- [단계 4-3] 입력변수 선택

[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.

추정된 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.

$$E_1 \geq E^* \quad (2)$$

단, E^* 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 종료판정을 사용하지 않고, 5층으로 제한하였다.

2.2 하중계수를 가진 목적함수

본 논문에서는 근사화 및 일반화의 상호 균형과 의존능력을 가진 합성 목적함수(성능지수)를 이용하여 상호 연계된 최적 모델 구조를 생성한다[9]. 이를 위해 학습 및 테스트 데이터의 성능 결과에 대한 하중계수를 가진 목적함수인 합성 성능지수로써 식 (3)과 같이 정의한다.

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

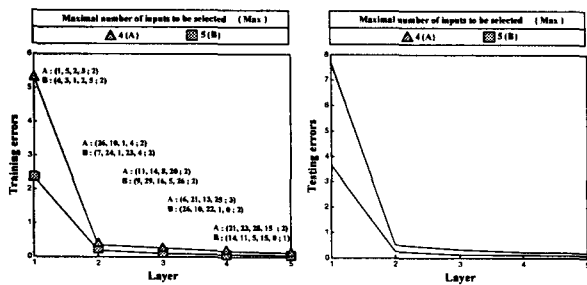
3. 시뮬레이션 및 결과 고찰

대기상 질소산화물(NOx)은 주로 연소공정을 통해 배출되며 산업공정, 발전소, 자동차 배기가스 그리고 난방에 의해서 발생하는데, 그 중 대기에서 문제가 될 만큼 존재하는 것들은 N2O, NO 및 NO2이고 이들 물질을 대기오염 측면에서 NOx(질소 산화물)로 일컫는다.

NOx의 저감방법으로는 발생원으로부터 NOx의 배출을 억제하는 방법과 배출가스중의 NOx를 제거하는 방법을 들 수 있는데, 최근에는 Gas Turbine에 사용하는 입출력 변수 중에 입력으로 Tamb, COM, LPT, Pcd, Texh 등의 주변 환경 변수를 조절하며 출력으로는 Parts per million-volume dry 단위의 NOx양으로 배출 메카니즘을 모델링하고, 최적 모델을 구축하여 시뮬레이터를 구성함으로써 NOx를 효과적으로 관리하는 방법이 연구되고 있다[8].

본 논문에서는 이러한 입출력 변수들을 이용하여 제안한 모델의 성능을 테스트하기 위해서 전체 260개의 데이터를 랜덤하게 130개씩 선택하여 학습데이터와 테스트 데이터로 사용하여 모델의 근사화와 일반화를 얻는다.

그림 2는 하중 계수 값 $\theta=0.5$ 일 때, 출력 \hat{y} 의 성능지수를 나타낸 것이다. 그림에서 \blacktriangle 4(A)는 4 입력이고, \blacksquare 5(B)는 5입력을 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 근사화와 일반화 능력이 뚜렷하게 향상되는 것을 알 수 있다.

그림 3은 하중계수 값이 $\theta=0.5$ 이고 최대입력이 5인 경우에 3층에서 네트워크를 나타낸 것이다.

그림 3에서 보여준 노드는 다음과 같이 표현된다. “ $\frac{PN_n}{NI \quad T}$ ” 여기서, ‘PNn’는 각 층에서 n번째

PN 노드를 나타낸 것이고, ‘NI’는 해당 노드로 들어오는 입력의 수를, ‘T’는 해당 노드에 사용된 차수를 나타낸 것이다.

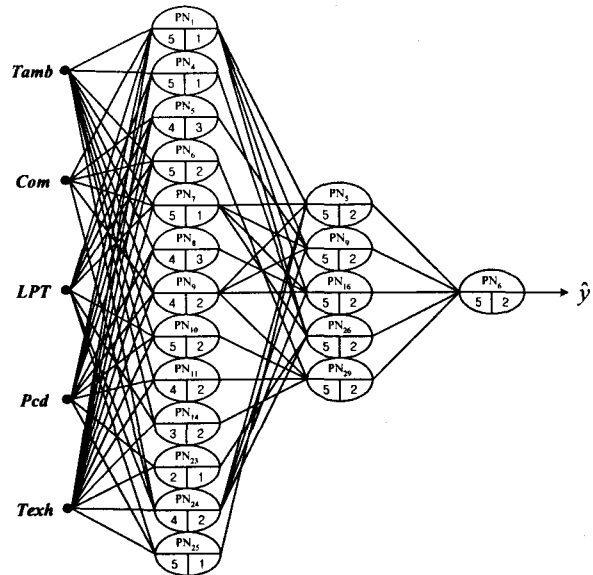
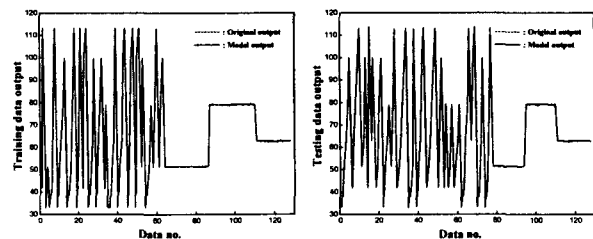


그림 3 GA-based PNN의 최적 구조 ($\theta=0.5$)

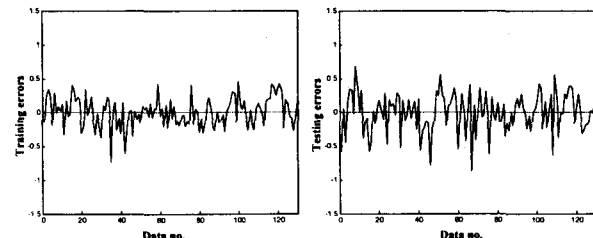
그림 4는 최적 구조일 때 모델 출력과 실제 출력의 차이를 그래프로 나타내고 있다.



(a) Training output (b) Testing output

그림 4 실제 출력과 모델 출력의 비교

그림 5는 학습 데이터와 테스트 데이터의 오차를 나타낸 것이다.



(a) Training error (b) Testing error

그림 5 GA-based PNN의 오차곡선

표 2는 가스 터빈 발전소의 NOx 배출 공정 데이터를 사용하여 성능 지수를 식 (3)에 의해 계산된 값으로 기존의 모델과 본 논문에서 제시

한 유전자 알고리즘을 이용한 PNN 모델과의 성능비교를 한 것이다.

표 8 기존 모델방법과의 동정 오차비교

Model			Performance Index		
			PI	PI _s	EPI _s
Regression model				17.68	19.23
Ahn's model[10]	FNN		5.835		
	AIM		8.420		
FNN[11]	Simplified	$\theta=0.4$		6.269	8.778
	Linear	$\theta=0.2$		3.725	5.291
Multi-FNN[12]	Linear	$\theta=0.75$		0.720	2.025
Our model		$\theta=0.5$		0.0433	0.0328

4. 결론

본 논문에서는 유전자 알고리즘을 이용하여 다항식 뉴럴 네트워크의 최적 구조를 설계하였다. 기존의 PNN 알고리즘은 층을 구성하는 노드의 입력변수의 수와 다항식 차수 그리고 고정된 입력변수 선택을 설계자의 경험에 의해서 결정함으로써 객관적이고 주어진 데이터의 특성에 맞는 최적의 모델을 구축하지 못하였고, 또한 최적의 모델을 구축하기 위해서 많은 시행착오를 필요로 하였다. 이러한 기존의 PNN 모델의 구조의 한계를 극복하기 위하여 최적화 방법인 유전자 알고리즘을 이용하여 노드의 입력변수의 수와 그에 해당하는 입력변수 그리고 다항식 차수 등의 파라미터를 최적 선택하여 모델 구축에 있어서 유연성과 정확성을 가지며 더욱 정확한 예측 능력을 가진 최적화된 PNN 모델구조를 구축하였다. 이러한 최적화된 GA-based PNN 구조를 비선형 공정에 적용하기 위하여 가스 터빈 발전소의 NOx 배출 공정 데이터를 통하여 성능을 확인하였다. 시뮬레이션을 통해서 알 수 있듯이 제안한 모델은 유전자 알고리즘을 이용한 각 노드들의 유기적인 조적으로부터 기존의 다른 모델들보다 성능이 우수함을 알 수 있으며, 데이터 특성이 비선형성이 강하거나 또는 제한된 데이터의 수 일지라도 구조적 및 파라미터 동정에 의한 효율적인 최적 다항식 뉴럴 네트워크를 구축할 수 있음을 보였다.

감사의 글

본 연구는 산업자원부의 지원에 의하여 기초전력연구원 (R-2004-B-274) 주관으로 수행된 과제임.

5. 참고문헌

[1] A. G. Ivahnenko, "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. Ivakhnenho, "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.

[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] G. Vachtsevanos, V. Ramani, and T. W. Hwang, "Prediction of Gas Turbine NOx Emissions using Polynomial Neural Network", Technical Report, Georgia Institute of Technology, Atlanta, 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.

[10] T.-C. Ahn, S.-K. Oh, A thesis of emission pattern model about the atmosphere pollution material of a power plant, Electrical Engineering & Science Research Institute, Korea, 1997.

[11] S.-K. Oh, W. Pedrycz and H.-S. Park, "Hybrid Identification in Fuzzy-Neural Networks", Fuzzy Sets & Systems, 2002.

[12] S.-K. Oh, W. Pedrycz and H.-S. Park, "Rule-based Multi-FNN Identification with the Aid of Evolutionary Fuzzy Granulation", Journal of Knowledge-Based Systems, 2002.