

GA기반 다항식 뉴럴네트워크를 이용한 비선형 모델링

김동원, 노석범, 이동윤, 오성권
원광대학교 공과대학 제어계측공학과

Nonlinear modeling by means of Ga based Polynomial Neural Networks

Dong-Won Kim, Seok-Beom Roh, Dong-Yoon Lee, and Sung-Kwun Oh
School of Electrical and Electronic Engineering, Wonkwang Univ.

Abstract - In this paper, Polynomial Neural Networks(PNN) is proposed to overcome some problems, such as the conflict between overfitting and good generation, and low reliability and to control nonlinearity and unknown parameter of complex system. PNN structure is consisted of layers and nodes like conventional neural networks but is not fixed and can be generated according to the system environments. The performances depend on two factors, number of inputs and order of polynomials in each node directly. In most cases these factors are decided by the trial and error of designer so optimization is needed in deciding procedure of the factors. Evolutionary algorithm is applied to decide the factors in PNN. The study is illustrated with the aid of representative time series data for gas furnace process used widely for performance comparison, and shows the designed PNN architecture with evolutionary algorithm.

항식과 다변수 입력에 의하여 차수 및 입력변수의 선택과 알고리즘의 최종종료판정 기준에 의하여, 파라미터 동정에 의한 모델을 선정하게 되고, 계층적 모델추정에 의하여 최종적으로 최적의 모델추정식을 얻는다. 입력력 데이터 $(x_i, y_i) = (x_{1i}, x_{2i}, \dots, x_{ni}, y_i), i=1, 2, \dots, n$ 가지 주어져 있다. PNN알고리즘은 이 데이터의 입력력 관계가 다음의 종속관계 f 를 만족한다고 생각한다.

$$y = f(x_1, x_2, \dots, x_n) \quad (1)$$

관계식 f 의 추정 모델 \hat{f} 를 다항식으로 판정하고, 출력 y 의 추정치 \hat{y} 는 다음의 관계식

$$\hat{y} = \hat{f}(x_1, x_2, \dots, x_n) = c_0 + \sum_{k_1} c_{k_1} x_{k_1} + \sum_{k_1, k_2} c_{k_1 k_2} x_{k_1} x_{k_2} + \sum_{k_1, k_2, k_3} c_{k_1 k_2 k_3} x_{k_1} x_{k_2} x_{k_3} + \dots \quad (2)$$

으로 표시된다. 단, c_k 는 계수이다. 연속 세대별 계층 구조에 의한 PNN알고리즘의 전체 개요도를 그림 1에 나타내었다.

1. 서 론

비선형문제, 시스템 입력변수 수의 선택, 모델의 높은 근사화와 일반화 능력의 탐색등의 문제점을 해결하고 비선형 공정의 복잡성과 불확실성을 제어하기 위해 다항식 뉴럴네트워크(Polynomial Neural Networks: PNN)가 제안되었다. 제안된 PNN구조는 노드의 입력변수와 차수가 모델 성능과 직접적으로 연관된다. 대부분의 경우, 이러한 요소들은 설계자의 시행착오에 의하여 결정된다. 따라서 최적의 성능을 얻기 위하여 각 노드의 부분표현식을 구성하는 입력변수와 차수를 결정하는 객관적이고 신뢰성 높은 새로운 자동 선택방법이 필요하다. 본 논문에서는 부분표현식을 구성하는 입력변수와 차수의 최적화를 위해 유전 알고리즘을 이용한다. 유전 알고리즘은 자연세계의 진화 현상에 기반한 계산 모델이다. 진화 알고리즘은 풀고자 하는 문제에 대한 가능한 해들을 정해진 형태의 자료구조로 표현한 다음 이들을 점차적으로 변형함으로써 점점 더 좋은 해들을 생성한다. 제안된 모델의 성능평가를 위해, 앞서 연구된 다양한 퍼지 및 지능형 모델링방법들에 광범위하게 적용된 가스로공정의 시계열데이터를 이용하여 결과를 고찰하고 또한 비교해석이 수반된다.

2. 다항식 뉴럴네트워크

GMDH 알고리즘에 기초한 PNN 알고리즘은 고계다

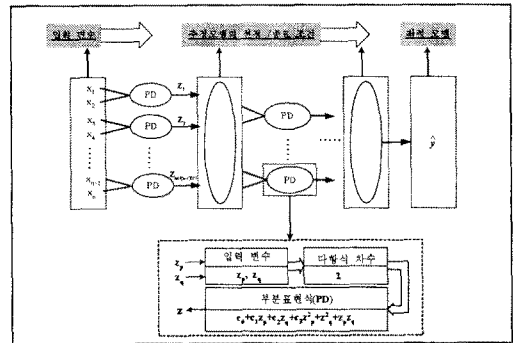


그림 1. PNN 구조의 전체 개요도

PNN 알고리즘의 단계를 살펴보면 다음과 같다.

- [단계 1] 입력변수 설정
출력변수 y 에 관련된 n 개의 입력변수를 설정한다.
- [단계 2] 데이터 분할
전체 데이터를 학습용 데이터와 평가용 데이터로 분할한다.
- [단계 3] PNN 구조 선택
기본 PNN 구조 및 변형된 PNN 구조를 선택한다.
- [단계 4] 노드의 차수 및 입력변수 수 선택
 n 개의 입력변수 중 노드의 입력변수와 차수선택
- [단계 5] 매개변수 추정
선택된 입력변수에 대한 다항식의 차수를 결정한다.
- [단계 6] 평가후 노드의 선택
평가용 데이터를 대입하고 2승 오차를 계산한다.
- [단계 7] 종료 판정
원하는 세대까지 알고리즘을 실행한 후 종료한다.
- [단계 8] 새로운 입력변수 선택

$X_{1j} = Z_{1j}, X_{2j} = Z_{2j}, \dots, X_{wj} = Z_{wj}$ 가 되어 새로운 입력력 데이터를 구성한다.

3. 진화 알고리즘

진화 알고리즘(evolutionary algorithm) 또는 유전 알고리즘(genetic algorithm: GA)은 자연세계의 진화 현상에 기반한 계산 모델이다. 진화 알고리즘은 풀고자 하는 문제에 대한 가능한 해들을 정해진 형태의 자료 구조로 표현한 다음 이들을 점차적으로 변형함으로써 점점 더 좋은 해들을 생성한다. 각각의 가능한 해를 하나의 개체(individual)로 보며 이들의 집합을 개체군(population)이라 한다. 하나의 개체는 보통 한 개 또는 여러개의 염색체(chromosomes)로 구성되며 염색체를 변형하는 연산자들을 유전 연산자(genetic operators)라 한다. 진화 알고리즘은 탐색, 최적화 및 기계학습의 도구로 많이 사용된다. 구조가 단순하고 방법이 일반적이어서 유전 알고리즘의 응용범위가 상당히 넓은 것이 특징이다. 유전 알고리즘이 기존의 탐색 또는 최적화 방법과 다른 점은 다음과 같다.

- ▶ 유전 알고리즘은 해공간(solution space)에서 단일 해를 사용하는 것이 아니라 해집단(population of solution space)을 이용하는 탐색방법이다.
- ▶ GA에서의 탐색은 확률적 연산자를 사용하여 수행되며, global한 최적해의 발견을 가능케 한다.
- ▶ 탐색공간에 대해 최적화 함수 정보(연속성이나 미분 가능성 등)의 제약이 받지 않는다.

4. 유전자 알고리즘과 PNN의 결합

PNN 알고리즘은 뉴럴네트워크의 구조와 유사하지만 층과 노드가 미리 결정되어 고정되지 않고 새롭게 생성될 수 있으며 각 노드가 1차, 2차, 변형된 2차 다항식과 같은 여러 가지 형태의 고계다항식을 사용하고, 다변수 입력으로 구성되어 있다. 또한 변수 선택이나 모델구조 결정에서 설계자의 선입견과 추측을 배제하고 객관적인 모델구축을 가능하게 되었다. 시스템의 환경에 맞게 적응하며 각층의 입력변수 수와 차수를 변형하면서 시스템을 효율적으로 표현하였다. 그러나 최적의 추정모델은 수 세대를 거친 계층의 부분표현식에 의하여 얻어지며, 각 계층의 부분표현식은 차수와 입력변수의 형태에 따라 설계가 수행됨으로 모델 성능과 직접적으로 연관된다. 따라서 최적의 성능을 얻기 위하여 각 노드의 부분표현식을 구성하는 입력변수와 차수를 결정하는 객관적이고 신뢰성 높은 새로운 자동 선택방법이 필요하다.

유전자 알고리즘은 자연 선택과 유전학에 기반한 확률적 탐색방법으로써 생산, 교차, 돌연변이의 과정을 수행하는 탐색법으로 비선형 최적화 이론으로 탁월한 성능을 발휘하고 있다. 특히 탐색영역이 넓어 광범위한 영역에서 방대한 지역 극소를 제거하고 반면에 모델의 성능지수가 최소가 되는 전역 극소 영역을 찾는다.

본 논문에서는 유전자 알고리즘으로 최적의 입력변수와 차수를 동조하여 PNN을 구현함으로써 기존의 모델구조보다 최적화된 모델을 구성할 수 있다.

5. 시뮬레이션 및 결과고찰

본 논문에서 제안한 진화론적 설계방법에 따라 PNN을 설계하여 가스로 공정[2]에 적용하여 모델의 타당성 및 유용성을 평가하였다. 모델의 성능을 평가하기 위해 전체 데이터 집합을 이용하여 학습과 테스트 데이터로 분할하였다.

입력변수 : $u(t-3), u(t-2), u(t-1), y(t-3), y(t-2), y(t-1)$ 로 입력 변수를 설정한다.

출력 변수 : $y(t)$ 로 출력 변수를 설정한다.

여기서, $u(t)$ 는 가스 흐름을이고 $y(t)$ 는 연소된 이산화탄소의 농도이다. 총 296개의 데이터 중 처음 148개

의 데이터는 학습에 이용하고 나머지 148개의 데이터는 테스트에 이용한다. 유전자 알고리즘을 이용하여 PNN의 각 노드의 입력변수와 차수를 결정하여 유전자 알고리즘의 하중값을 가진 목적함수에 따라 모델이 결정된다. 그럼으로써 일반화된 PNN 모델의 기본형태와 변형된 형태가 자동 조절된다.

[단계 4]에서 선택한 입력변수 수와 후반부 다항식에 의해 표현된 부분표현식으로부터 분할한 학습용 데이터를 사용하여 최소자승법에 의해 식 (3)을 최소화하는 최적의 파라미터를 구한다.

$$E = \frac{1}{2} \sum_{i=0}^N (y_i - z_{ik})^2 \quad (3)$$

학습 데이터를 사용하여 얻은 모델로부터 테스트 데이터를 사용하여 동정오차 E 를 계산하여 모델을 평가한다. 진화이론에 의해 설계된 PNN은 전체 층수를 3층으로 제한하여 알고리즘을 수행한다.

진화설계로 구축된 PNN의 성능지수를 아래의 그림에서 보였으며, 학습 데이터만을 고려한 기존모델 방법의 동정오차와 비교하여 앞선 다른 연구에서보다도 더 높은 정확도를 가짐을 보이고, 학습 데이터뿐만 아니라 평가용 데이터도 함께 고려하여 이에 따른 결과도 더 높은 정밀도를 가짐을 보인다. 유전자 알고리즘의 구성요소로 개체표현방법은 2진 스트리밍을 사용했으며, 유전 연산자로 선택(selection)연산자는 roulette-wheel방식, 교배(crossover)연산은 one-point crossover 연산을 사용한다. 돌연변이 연산은 선택된 비트를 반전시키는 invert방식을 사용한다. 세대는 2세대, 군집은 5개체까지 진행시켰으며, 각 문자열의 길이는 5 bit를, 교배율은 0.6, 돌연변이 발생율은 0.35를 PNN 각 층에서 최적의 입력변수와 차수를 탐색하기 위해 사용한다.

아래의 그림 2-4는 PNN의 입력변수와 차수를 유전자 알고리즘으로 동조하여 얻은 성능지수의 최적화 과정을 나타낸 것이다. 그림 2는 MPI의 최적화 과정을 보여주며, 그림 3은 PI, 그림 4는 EPI의 최적화 과정을 나타낸다. 그림 5에서는 각각의 성능지수가 PNN의 층이 증가함에 따라 변화하는 것을 보여준다. 그림 2는 PNN의 각 층에서 생성된 MPI를 그래프로 나타낸 것으로 $MPI = (PI + EPI) / 2$ 를 의미한다. 3층에서 0.066으로 가장 좋은 값을 보여준다. 그림 3은 PNN의 각 층에서 생성된 PI를 그래프로 나타낸 것으로 3층에서 0.013로 가장 좋은 값을 보여준다. 그림 4는 PNN의 각 층에서 생성된 EPI를 그래프로 나타낸 것으로 3층에서 0.119로 가장 좋은 값을 보여준다.

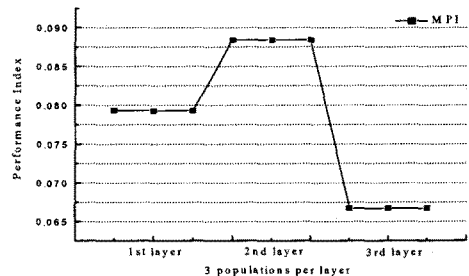


그림 2. 유전자 알고리즘에 의한 PNN의 성능지수(MPI)의 최적화 과정

표 1은 기존의 모델링 방법들과 제안된 모델링 방법의 성능지수를 보여준다. 표에서 알 수 있듯이 제안된 모델은 시스템의 비선형성, 입력변수에 따른 복잡성 등 데이터의 특징을 모델에 반영하여 모델구조 및 입력변수의 수 그리고 시스템 방정식의 차수를 최적화 알고리즘인 진화(유전자) 알고리즘에 의하여 선택하여 최적의 모델을 선정하므로 모델의 정밀도가 높다. 그러므로 기존의 모델들 보다 훨씬 좋은 학습 성능결과를 얻었다. 본

논문에서 제안한 PNN 구조는 기존 모델들의 학습 성능 오차보다 우수할 뿐만 아니라, 테스트 데이터를 고려한 평가성능 즉 예측성능 면에서도 우수한 특성을 보임을 알 수 있다.

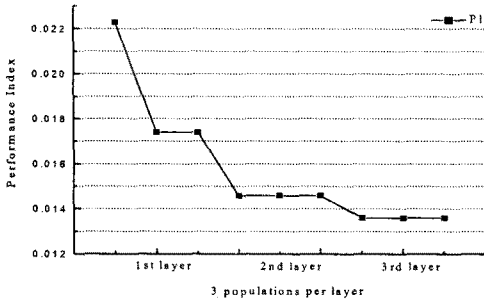


그림 3. 유전자 알고리즘에 의한 PNN의 성능지수(PI)의 최적화 과정

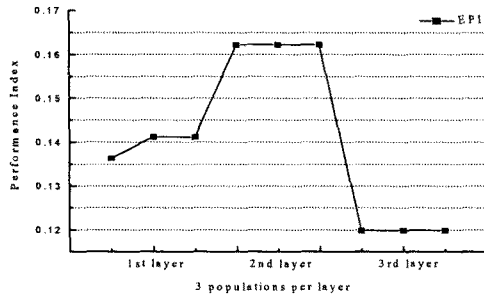


그림 4. 유전자 알고리즘에 의한 PNN의 성능지수(EPI)의 최적화 과정

표 1. 기존 모델방법과의 동정 오차비교

Model	MSE		
	PI	PI	EPI
Tong's model[3]	0.469		
Sugeno and Yasukawa's model[4]	0.190		
Xu's model[5]	0.328		
Pedrycz's model[7]	0.320		
Oh and Pedrycz's model[8]	0.123	0.020	0.271
Kim, et al.'s model[6]		0.034	0.244
Lin and Cunningham's model[9]		0.071	0.261
Our model (Evolutionary algorithm+PNN)		0.013	0.119

진화알고리즘에 의하여 동조된 PNN의 입력변수 수와 차수는 5 입력 Type 1이며 이로 인하여 PNN의 성능지수인 PI와 EPI는 각각 0.013과 0.119를 얻었다.

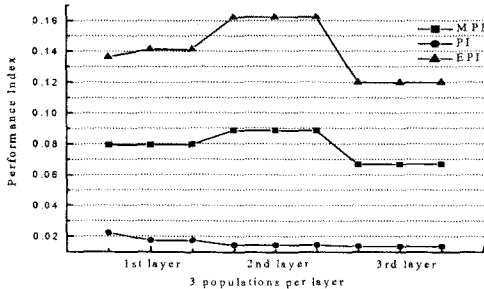


그림 5. 유전자 알고리즘에 의한 PNN의 전체 성능지수의 최적화 과정

6. 결론

본 논문에서는 데이터 수가 적거나 혹은 복잡한 비선형 요소가 많은 시스템의 체계적이고 효율적인 최적 모델을 얻기 위하여 다항식 뉴럴네트워크(Polynomial Neural Networks: PNN) 구조의 최적 설계와 관련하여 진화론적 설계방법을 제안하였다. 시뮬레이션에서 알 수 있듯이 본 논문에서 제안된 모델이 기존의 다른 지능형 모델보다 학습과 테스트 성능면에서 아주 우수한 것은 물론 다음과 같은 특징들이 있음을 알 수 있었다.

- ▶ PNN의 구조는 일반적인 다층 퍼셉트론 구조의 경우와 달리 미리 정해져 있지 않고 진화알고리즘에 의하여 최적화 되는 동안 모델 스스로 노드의 선택과 제거를 통해 네트워크 구조를 생성한다.
- ▶ PNN 모델은 근사화를 위한 overfitting과 일반화 사이의 충돌을 감소시킬 수 있다.

감사의 글

본 연구는 한국과학재단 목적기초연구 (과제번호: R02-2000-00284) 지원으로 수행되었음.

[참고 문헌]

- [1] A.G. Ivahnenko, "Polynomial theory of complex systems", *IEEE Trans. On Systems, Man and Cybernetics*, Vol. SMC-1, pp. 364-378, 1971.
- [2] G.E.P. Box and F.M. Jenkins, "Time Series Analysis : Forecasting and Control", 2nd ed. Holden-day, 1976
- [3] R. M. Tong, "The evaluation of fuzzy models derived from experimental data, *Fuzzy Sets and Systems*, Vol.13, pp.1-12, 1980.
- [4] M. Sugeno, and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling", *IEEE Trans. Fuzzy Systems*, 1, (1), (1993), 7-31
- [5] C. W. Xu., and Y. Zailu, "Fuzzy model identification self-learning for dynamic system, *IEEE Trans on Systems, Man, Cybernetics*, Vol. SMC-17, No.4, pp.683-689, 1987.
- [6] E. Kim, H. LEE, M. Park, and M. Park, "A simple identified Sugeno-type fuzzy model via double clustering, *Information Sciences*, 110, pp.25-39, 1998.
- [7] W. Pedrycz, "An identification algorithm in fuzzy relational system, *Fuzzy Sets and Systems*, Vol. 13, pp.153-167, 1984.
- [8] 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, issue 2, pp. 205-230, Jul. 2000.
- [9] Y. Lin, G. A. Cunningham III, "A new approach to fuzzy-neural modeling", *IEEE Trans. Fuzzy Systems* 3, (2), (1995), 190-197.
- [10] S.-K. Oh, W. Pedrycz and D.-W. Kim, "Hybrid Fuzzy Polynomial Neural Networks", *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*(Submitted)
- [11] 오성권, 김동원, "적용 다항식 뉴로-퍼지 네트워크 구조에 관한 연구", *Trans. KIEE*, Vol. 50D, No. 9, pp. 430-438, Sep., 2001.
- [12] 오성권, 윤기찬, 김현기, "유전자 알고리즘과 합성 성능지수에 의한 최적 퍼지-뉴럴 네트워크 구조의 설계", *Journal of Control, Automation and Systems Engineering*, Vol. 6, No. 3, March 2000.