

# 유전자 알고리즘에 의한 자기구성 다항식 뉴럴네트워크 최적 설계

박호성\* · 박병준\*\* · 오성권\*\*\*

Optimal Design of Self-Organizing Polynomial Neural Networks By  
Means of Genetic Algorithms

*Park, Ho-Sung\* · Park, Byoung-Jun\*\* · Oh, Sung-Kwon\*\*\**

## Abstract

In this paper, we proposed optimal design of Self-Organizing Polynomial Neural Networks (SOPNN) by means of Genetic algorithms(GAs). The proposed SOPNN is based on the Group Method of Data Handling method. The SOPNN consists of layers and nodes like conventional neural networks but is not fixed and can be generated according to the system environments. The each node of SOPNN used several types of high-order polynomial such as linear, quadratic, and modified quadratic, and is connected with various kinds of multi-variable inputs.

In this paper, GAs is applied to improve the performance with optimal input variables and order.

The study is illustrated with the aid of representative time series data for gas furnace process used widely for performance comparison, and shows the GA-based SOPNN architecture can produce the model with higher accuracy than pervious other works. And performance index related to approximation and prediction capabilities of model is evaluated and also discussed.

## 1. 서론

효율적인 제어메커니즘의 설계와 알려지지 않은 시스템의 입출력데이터로 시스템을 파악하는 것은 피할 수 없는 어려운 문제이며 시스템의 비선형성

과 입출력관계에 대한 제한된 양의 지식과 정보, 시스템 입력변수 수의 선택, 모델의 높은 근사화와 일반화 능력의 탐색 등의 문제에 직면한다. 따라서 일반적인 수학적 방법의 적용보다는 진보된 지능 기반 연산기술에 많은 관심이 집중되고 있다. 그 중에서 새로운 모델링 기법으로 입력과 출력사이의 비선형관계를 표현함에 있어 선택된 입력변수와 차수를 이용하는 자기구성 다항식 뉴럴네트워

\* 원광대학교 제어계측공학과 박사과정

\*\* 원광대학교 전기공학과 박사과정

\*\*\* 원광대학교 전기전자및정보공학부 부교수

크(Self-organizing Polynomial Neural Networks: SOPNN)[1-4]가 있다.

SOPNN 알고리즘은 입력변수의 선택과 부분표현식을 정의하여 시스템 방정식을 예측하므로 입출력 관계나 모델의 함수형이 특정하지 않은 비선형 관계의 경우에 변수와 차수로 정의된 부분표현식을 계층적으로 조합하여 비선형 모델 추정식을 얻을 수 있다. 각 계층은 부분표현식에 의한 회귀분석을 적용함으로써, 입출력 관계의 비선형 함수를 용이하게 얻을 수 있고, 매우 적은 데이터로도 복잡하고 다변수인 비선형계의 동정과 예측이 가능하다. 이는 기존의 신경망 알고리즘보다 더 적은 노드수와 유연한 적응특성을 가지며 그 기본 알고리즘은 GMDH[5,6]을 기반으로 하고 있다.

그러나 이와 같은 이론적인 발전과 많은 실제적인 성공에도 불구하고 비선형 시스템에 적용하여 바람직한 결과를 얻기 위해서는 구조적인 최적화 과정이 필요하다. SOPNN구조에서 노드의 입력변수와 차수는 다항식에서 사용되는 변수의 수와 차수에 종속되어 모델 성능과 직접적으로 연관됨으로 모델구조의 성능을 향상시키기 위한 중요한 요소이다. 대부분의 경우, 이러한 요소들은 설계자의 경험에 의하여 결정된다. 따라서 최적의 성능을 얻기 위하여 각 노드의 부분표현식을 구성하는 입력변수와 차수를 결정하는 객관적이고 신뢰성 높은 새로운 자동 선택방법이 필요하다. 따라서 본 논문에서는 부분표현식을 구성하는 입력변수와 차수를 최적화하기 위해 유전 알고리즘을 이용한다.

1970년대 초 J. Holland에 의해 제시된 유전 알고리즘[7]은 자연계에 존재하는 적자생존의 원리와 유전학에 기초하여 집단내 개체들을 모의 진화시킴으로써 전역해(global minima)를 찾아가는 최적화 기법이다. 유전 알고리즘이 기존의 최적화 이론들과 근본적으로 다른 점은 목적함수의 도함수를 이용하지 않으며 다수의 변수들을 하나의 문자열로 나타낸 집단을 사용함으로써 병렬연산을 통한

최적화가 가능하다는 것이다. 또 확률적인 방법에 의해 전역탐색을 수행함으로써 부분극소점(local minima)에 빠질 가능성이 적다.

본 논문에서는 자기구성 다항식 뉴럴네트워크(SOPNN)의 구조와 알고리즘을 설명하고, SOPNN 구조를 최적화하기 위해서 유전 알고리즘에 의한 SOPNN 알고리즘의 최적구조설계를 연구하였다. 제안된 모델의 성능평가를 위해, 가스로 공정의 시계열데이터[8]를 모델링하여 결과를 고찰하고 기존 모델과 성능 비교해석을 수반한다.

## 2. 본 론

### 2.1 자기구성 다항식 뉴럴 네트워크

GMDH 알고리즘에 기초한 SOPNN 알고리즘은 고계다항식과 다변수 입력에 의하여 차수 및 입력변수의 선택과 알고리즘의 최종종료판정 기준에 의하여, 파라미터 동정에 의한 모델을 선정하게 되고, 계층적 모델추정에 의하여 최종적으로 최적의 모델추정식을 얻는다. 입출력 데이터  $(x_i, y_i) = (x_{i1}, x_{i2}, \dots, x_{in}, y_i)$ ,  $i=1, 2, \dots, n$ 까지 주어져 있다. SOPNN 알고리즘은 이 데이터의 입출력 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 관계식으로 표시된다.

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

단,  $\alpha_k$ 는 계수이다. SOPNN에서 추정모델  $\hat{f}$ 를 얻기 위해, 먼저 부분표현식을 구성한다. 부분표현식으로 표 1과 같은 회귀다항식을 사용한다. 다음에, 선형회귀분석으로 부분표현식의 계수를 구하고, 중간변수  $z_k$ 를 구한다. 중간변수에 의한 입출력 데이

터를 구성하여, 다음 계층에서도 이와같은 형태의 연산을 되풀이하며, 알고리즘의 종료 판정식을 만족할 때까지 계속 반복된다. 알고리즘을 종료한 후 얻어지는 부분표현식은 계층적으로 조합하고, 최종 추정모델  $\hat{f}$  를 얻는다.

다음은 SOPNN 알고리즘 설계 절차이다.

**[단계 1] 입력변수 설정**

출력변수  $y$ 에 관련된  $n$ 개의 입력변수를 설정한다. 필요하면 데이터를 정규화 한다.

이 입력변수를  $x_1, x_2, \dots, x_n$ 이라고 하자.

**[단계 2] 데이터 분할**

$N$ 개의 입출력 데이터  $(x_i, y_i) = (x_{1i}, x_{2i}, \dots, x_{ni}, y_i)$ ,  $i=1, 2, \dots, N$ 을 학습용 데이터와 평가용 데이터로 분할한다. 학습용 데이터는 모델을 동정하기 위해 사용하는 데이터이고, 평가용 데이터는 모델을 평가하기 위해 사용된다. 학습용 데이터와 평가용 데이터의 개수를 각각  $N_t$ 개와  $N_c$ 개로 한다. 여기서,  $N = N_t + N_c$ 이다.

데이터를 학습용과 평가용으로 분할하는 방법으로는, 입출력 데이터의 최초 데이터로부터 학습용 데이터와 평가용 데이터로 상호 분배하는 방법, 데이터 번호의 난수를 이용하여 분할하는 방법, 출력 데이터의 분산이 큰 데이터를 학습 데이터로, 작은 쪽을 평가용 데이터로 분할하는 방법, 그 역의 방법 등이 있다.

**[단계 3] SOPNN 구조 선택**

SOPNN 구조는 1 세대와 그 이후세대의 부분표현식을 구성하는 방법에 따라 모델의 구조가 정해지는데, 1 세대와 그 이후 세대가 각각 서로 동일 입력변수 수를 가지는 경우와 서로 다른 입력변수 수를 가지는 경우의 두 구조로 분류되며 이를 기본 SOPNN 구조, 변형된 SOPNN 구조라 한다. 표 2에서 모델구조를 나타내었다.

(a) 기본 SOPNN 구조 - 동일 입력변수 수

Case 1. 동일 차수를 가지는 구조.

Case 2. 서로 다른 차수 혹은 변형된 형태의 차수를 가지는 구조.

(b) 변형된 SOPNN 구조 - 서로 다른 입력변수 수

Case 1. 동일 차수를 가지는 구조.

Case 2. 서로 다른 차수 혹은 변형된 형태의 차수를 가지는 구조.

**[단계 4] 노드의 차수 및 입력변수 수 선택**

$n$ 개의 입력변수  $x_1, x_2, \dots, x_n$  중 노드의 입력변수를 선택한다. 입력변수의 개수에 따라 여러 가지 형태의 부분 표현식이 구성된다.  $r$ 개의 변수 선택 시 입력변수의 개수는  $k = n! / ((n-r)!r!)$ 가 되며  $n$ 은 전체 입력변수의 수이고,  $r$ 은 선택된 입력변수의 수이다. 노드의 차수 및 입력변수 수 선택은 시스템의 특성, 모델 설계방법, 비선형성 및 예측성능 등을 고려함으로써 제기된 문제에 대해 가장 좋은 선택구조를 결정하는데 도움을 준다. 표 1은 입력변수 수에 따른 부분 표현식의 형태를 나타낸 것이다.

입력변수	다항식 차수	다항식 구조
2 변수	Type 1	$c_0 + c_1x_1 + c_2x_2$
	Type 2	$c_0 + c_1x_1 + c_2x_2 + c_3x_1^2 + c_4x_2^2 + c_5x_1x_2$
	Type 3	$c_0 + c_1x_1 + c_2x_2 + c_3x_1x_2$

표 1. 입력변수 수에 따른 부분표현식(Partial Description : PD)의 형태

**[단계 5] 매개변수 추정**

$n$ 개의 입력변수  $x_1, x_2, \dots, x_n$ 에서 노드의 입력변수를 선택하고, 선택된 입력변수에 대한 다항식의 차수를 결정한다. 매개변수의 추정은 선형 회귀분석에 의해 학습용 데이터를 사용하여, 최소자승법에 의해 식 (3)을 최소화하는 최적의 파라미터  $c_i$  를 구한다.

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

여기서,  $N_i$ 는 데이터의 총수를 나타낸다.

**[단계 6] 평가 후 노드의 선택**

계수  $\omega, c_1, \dots, c_5$ 를 추정한 부분표현식에 평가용 데이터를 대입하고, 평가용 데이터에 대한 식 (3)의 2승 오차  $E$ 를 계산한다. 노드선택 방법은 다음과 같은 방법으로 구분할 수 있다.

얻어진  $n!/((n-r)!r!)$ 개의  $E_1, E_2, \dots, E_{n!/((n-r)!r!)}$ 에 대하여, 값이 작은 순으로 나란히 정렬하고 이들 중 값이 작은 순으로 다시  $w$ 개를 선택하여,  $E_1, E_2, \dots, E_{n!/((n-r)!r!)}$ 로 재배열하고,  $w$ 개를 제외한 나머지 값들은 버린다.

- $n!/((n-r)!r!) < w$ 일 경우 다음세대 노드의 개수는  $n!/((n-r)!r!)$ 가 됨
- $n!/((n-r)!r!) \geq w$ 일 경우 다음세대 노드의 개수는  $W$ 가 됨

각 세대의 노드수를 일정하게 정의할 수 있으므로 모델링 작업에서 오는 많은 계산량과 과대한 연산시간을 방지할 수 있는 장점이다.

**[단계 7] 종료 판정**

알고리즘을 종료하는 방법에는 다음의 방법이 사용된다. [단계 6]에서 얻어진  $E_1$ 이 다음의 부등식을 만족하는 경우에 알고리즘을 종료한다.

$$E_1 \geq E. \tag{4}$$

( $E$ 는 전세대의 2승 오차  $E_1$ 이다)

**[단계 8] 새로운 입력변수 선택**

식 (4)를 만족하지 않을 경우,  $x_{1i} = z_{1i}, x_{2i} = z_{2i}, \dots, x_{wi} = z_{wi}$ 가 되어 새로운 입력력 데이터를 구성하고, 단계 5로 다시 돌아간다. 이후, 단계 5부터 단계 8까지를 반복한다. 알고리즘이 종료할 경우, 2승 오차  $E$ 를 얻은 부분표현식에 전 세대에서의 중간변수를 대입하고, 동일한 조작을 되풀이하여 최종 추정모델  $\hat{f}$ 를 얻는다.

**2.2 유전자 알고리즘**

유전 알고리즘(GAs)은 자연세계의 진화 현상에 기반 한 계산 모델이다. 유전자 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 여러 개의 염색체로 구성되며 염색체를 변형하는 연산자들을 유전 연산자(genetic operators)라 한다. 유전자 알고리즘은 탐색, 최적화 및 기계학습의 도구로 많이 사용된다. 구조가 단순하고 방법이 일반적이어서 유전 알고리즘의 응용범위가 상당히 넓은 것이 특징이다. 최적화 문제에서 탁월한 성능을 발휘하는 유전자 알고리즘[7]은 기존의 방법들과는 달리 선형, 연속, 미분가능 등의 제한이 없기 때문에 다양한 분야에 별다른 제한 없이 적용할 수 있다는 장점을 가진다. 최근의 많은 탐색방법들은 한 번에 최적값을 찾아내지 않고, 반복을 통해서 최적지점에 접근해 가는 방법을 사용하고 있다. 유전자 알고리즘도 마찬가지로 적정 수준에 도달할 때까지 탐색과정을 반복한다. 제안한 모델의 최적화를 위하여 사용된 유전자 알고리즘은 반복을 통해서 최적지점에 접근해 가는 방법을 사용하고 탐색지점으로 실수가 주어지면 이를 2진수(binary)로 변환하게 된다. 변환된 각 비트는 염색체처럼 취급되고, 염색체가 모여 유전자를 이루며 개체(individual)라 불린다.

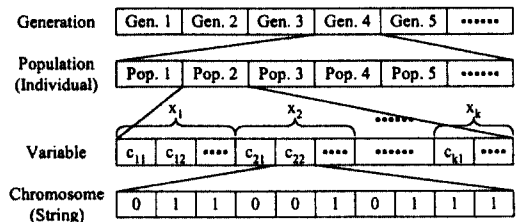


그림 1. GAs 데이터 구조의 계층 표현

즉, 각 염색체는 그림 1과 같이 모두 하나의 개체 안에 연결해 사용한다. 염색체(변수)는 사용자에게 의해서 bit로 정의하고, 개체가 탐색할 때 사용

되는 하나의 지점을 의미한다. 동시에 여러 점을 탐색하기 위해 개체를 동시에 여러 개 사용되는데 이를 군집이라 한다.

### 2.3 유전자 알고리즘을 이용한 자기구성 다항식 뉴럴 네트워크의 최적 설계

기존의 SOPNN 알고리즘[1-4]은 뉴럴네트워크의 구조와 유사하지만 층과 노드가 미리 결정되어 고정되지 않고 새롭게 생성될 수 있으며 각 노드가 1차, 2차, 변형된 2차 다항식과 같은 여러 가지 형태의 고계 다항식을 사용하고, 다변수 입력으로 구성되어 있다. 그러나 최적의 성능을 얻기 위하여 각 노드의 부분표현식을 구성하는 입력변수와 차수를 결정하는 객관적이고 신뢰성 높은 새로운 자동 선택방법이 필요하다.

본 논문에서는 유전자 알고리즘으로 최적의 입력변수와 차수를 동조하여 SOPNN을 구현함으로써 기존의 모델구조보다 최적화된 모델을 구성할 수 있다. 기존의 SOPNN 모델은 여러 가지 비선형의 복잡한 공정에서 두드러진 우수한 특성[1-4]을 보여왔다. 그러나 비선형 모델 추정식을 얻기 위해 설계자가 정의한 입력 변수 수로 생성될 수 있는 모든 경우의 수를 조합하여 모델 추정식을 얻었다. 이는 입력변수가 많아지면 그에 따라 네트워크가 상당히 복잡해 질 수 있을 뿐만 아니라 최적화를 얻기 위한 시간 또한 많이 걸린다[9,10]. 이것은 해의 수가 작을 때는 이들을 낱알이 열거하는 방식으로 찾을 수 있지만, 대단히 클 때는 현실적으로 불가능하게 된다. 따라서 유전자 알고리즘을 이용하여 자기구성 다항식 뉴럴네트워크의 최적 설계에 대한 이러한 문제점들의 해결책을 제시하는 것과 동시에 복잡한 다변수계의 시스템에 대하여 효율적인 입력변수와 차수 선택문제를 해결하고, 프로세스에 대한 충분한 지식이나 정보의 결핍에도 불구하고 불확실성에 대한 우수한 제어능력과 더불어 비선형성이 가미된 복잡한 시스템의

모델링에서 최적화된 모델구조생성으로부터 신뢰성 있는 근사화 능력과 일반화(예측) 능력을 보인다. 그림 2는 유전자 알고리즘에 의한 입력변수의 최적선택을 위한 염색체 코딩전략을 나타낸 것이다. 본 논문에서는 입력변수와 차수의 선택문제를 유전자의 염색체를 이용하여 해결하였다. 모든 변수를 이진 스트링으로 나타내는 방법으로 염색체의 길이는 <입력변수 선택용 구조 비트>+<차수 선택용 구조 비트>의 수와 일치한다. 그림 2에서 보여진 것과 같이, 유전 알고리즘을 이용하여 입력변수와 차수의 최적선택을 위한 염색체 코딩전략은 다음과 같다. 먼저 <차수 선택용 구조 비트>의 경우, 2진 스트링을 10진수로 변환하여 그 변환값에 따라 차수를 선택하고, <입력변수 선택용 구조 비트>인 경우, 염색체의 특정 위치의 비트가 1이 되면 그 입력변수가 자기구성 다항식 뉴럴네트워크의 부분표현식을 구성하는 변수에 포함됨을 의미하고, 0은 반대의 의미를 가진다.

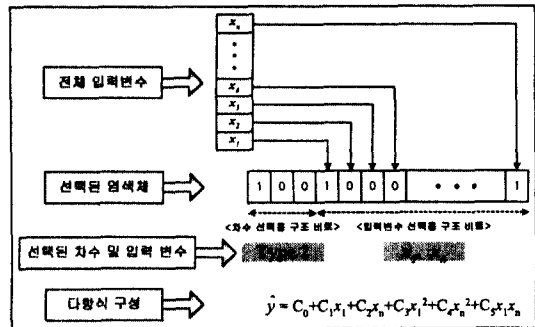


그림 2. 유전자 알고리즘에 의한 입력변수의 최적선택을 위한 염색체 코딩전략

본 논문에서 사용한 유전자 알고리즘에서 재생산은 Roulette-Wheel 방식을 사용하여 개체들을 선택한다. 교배는 한 점을 기준으로 교배가 이루어지는 일점 교배를 통하여 재결합되고 집단에 새로운 정보를 도입하면서 특정 정보를 잃지 않게 하기 위한 돌연변이 연산이 수행된다.

Maximum generation	100	100	100
Population size	60	60	60
Elite population size	20	20	20
bit length	3+6	3+20	3+20
Crossover rate	0.65		
Mutation rate	0.1		
Weighting factor	0.0~1.0		
Type (차수)	1~3		

표 2. SOPNN의 최적 설계를 위한 유전자 알고리즘의 매개변수와 설정치

이렇게 하여 SOPNN이 최적의 성능을 발휘할 때까지 새롭게 형성된 집단은 다시 평가되고 앞에서 수행한 일련의 연산 과정을 반복한다. 표 2는 본 논문에서 사용한 유전자 알고리즘의 정보를 나타낸 것이다.

## 2.4 하중값을 가진 목적함수

본 논문에서는 주어진 학습 데이터를 이용하여 SOPNN 모델을 구축하고 테스트 데이터를 사용하여 설계된 모델을 평가한다. 또한 모델평가의 척도인 근사화 및 일반화 능력의 상호 균형과 의존 능력을 가진 합성 목적 함수(성능지수)를 이용하여 상호 연계된 최적 모델 구조 생성에 의한 성능지수의 향상을 꾀하였다. 이를 위해 학습 및 테스트 데이터의 성능 결과에 대한 하중값을 가진 목적함수를 다음과 같이 정의한다.

$$f(PI, EPI) = \theta \times PI + (1 - \theta) \times E\_PI$$

$$fitness = \frac{1}{1 + f(PI, E\_PI)} \quad (5)$$

$PI$ 는 학습데이터에 대한 성능지수,  $E\_PI$ 는 테스트데이터에 대한 성능지수, 그리고  $\theta$ 는  $PI$ 와  $E\_PI$ 에 대한 하중계수를 나타내며, 목적함수는 모델의 성능향상이 최적이 되도록 하기 위해 SOPNN의 각 노드에 인가되는 입력변수가 유전자 알고리즘에 의해 자동 조절되며, 차수는 설계자에 의하여 고정된다.

식 (5)는 하중계수값 설정에 따라  $\alpha \in [0, 1]$ 에

대해서  $\theta = \alpha$ 이면 학습과 테스트 모두를 포함하고,  $\alpha$ 의 선택은 근사화와 일반화 능력 사이에서 상호 균형에 의한 최적화를 시도한다. 이것의 의미는 학습 데이터에 의한 SOPNN모델의 성능지수  $PI$ 를 구하고, 이 학습 데이터에 의해 생성된 SOPNN 모델에 테스트 데이터를 가하여 평가용 성능지수  $E\_PI$ 를 구한다. 구한  $PI$ 와  $E\_PI$ 에 하중계수  $\theta$ ,  $1 - \theta$ 를 곱한 다음 합에 의한 목적함수  $f(\cdot)$ 를 최적화시키는 방향으로 모델의 파라미터를 동조함으로써 SOPNN을 구축한다. 이것은 모델 자체의 근사화 능력과 일반화 능력을 상호 고려한 목적함수의 형태로 적절한 하중계수 설정에 따라 최적의 SOPNN을 구축할 수 있다.

## 3. 시뮬레이션 및 결과고찰

본 논문에서 제안한 SOPNN의 진화론적 설계의 유용성과 모델의 성능평가를 위해 Box와 Jenkin이 사용한 가스로 시계열 데이터[8]를 이용하여 입출력 데이터인 가스흐름율과 연소된 이산화탄소의 농도를 모델링 하였다. 가스로 공정 시계열 데이터에서 시스템 입력변수로는  $u(t-3)$ ,  $u(t-2)$ ,  $u(t-1)$ ,  $y(t-3)$ ,  $y(t-2)$ ,  $y(t-1)$ , 출력 변수로는  $y(t)$ 를 사용한다. 여기서  $u(t)$ 는 가스 흐름율이고,  $y(t)$ 는 연소된 이산화탄소의 농도를 나타낸다. 총 296개의 데이터 중 처음 148개의 데이터는 학습에 이용하고 나머지 148개의 데이터는 테스트에 이용한다.

전체 6개의 시스템입력 중 노드의 입력변수와 차수가 유전자 알고리즘에 의하여 염색체로 자동 선택한다. 이에 따라 여러 가지 형태의 부분표현식과 노드의 차수 구성된다. 분할된 학습용 데이터를 사용하여 최소자승법에 의한 SOPNN의 각 노드의 파라미터를 구하고, 여기서 추정된 계수  $\omega$ ,  $c_1$ ,  $c_2$ , ...,  $c_n$ 을 각각의 부분표현식에 대입한다. 평가용 데이터를 사용하여 동정오차(혹은 성능지수)를 구

하고, 하중값을 가진 목적함수에 의하여 최적의 입력변수들의 조합을 선정한다. 이 동정오차는 실제 시스템의 출력 데이터와 제안된 구조의 출력간의 평균자승오차(MSE)로 모델의 정밀도를 평가할 수 있으며 식 (6)으로 정의된다.

$$PI = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (6)$$

또한 다음 세대로의 입력을 위한 현 세대의 출력값은 먼저 에러가 작은 순으로 정렬을 한 다음 에러가 중복되는 염색체는 삭제를 하고 새로이 정렬을 하게 된다. 그럼으로써 다음세대로의 중복 입력을 피하고자 한다. 최종적으로 따라서 1 층의 출력값은 2 층의 입력값이 되어 새로운 입출력데이터를 구성하며, 종료판정기준에 따라 최대 3 층까지 동일한 조작을 반복하여 최종 추정모델을 얻는다.

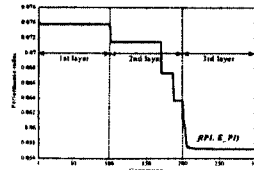
표 3은 유전자 알고리즘을 이용한 SOPNN을 설계한 성능결과를 나타낸 것이다.

$\theta$	1층		2층		3층	
	Type	$PI$	Type	$PI$	Type	$PI$
0.0	1	0.035	2	0.011	1	0.012
0.25	1	0.021	2	0.015	1	0.015
0.5	1	0.021	2	0.015	1	0.010
0.75	1	0.021	2	0.012	1	0.010
1.0	2	0.013	3	0.007	2	0.001

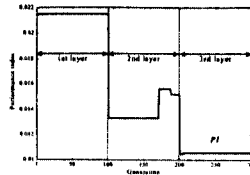
표 3. 유전자 알고리즘을 이용한 SOPNN 성능

표 3에서 알 수 있듯이  $\theta=0.5$ 일 때  $PI$ 는 0.010이고,  $E_{PI}$ 는 0.103으로 가장 에러값을 얻었다. 그림 3은 위의 경우를 유전자 알고리즘에 의한 SOPNN이 연속적인 세대에 의한 성능지수가 최적화되는 과정을 나타낸 것이다.

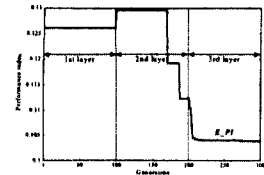
그림 4는 3층에서  $\theta=0.5$ 인 경우의 원 데이터 출력과 모델의 출력을 보여준 것이다.



(a)  $f(PI, E_{PI})$



(b)  $PI$



(c)  $E_{PI}$

그림 3. 유전자 알고리즘에 의한 SOPNN 성능지수 최적화 과정

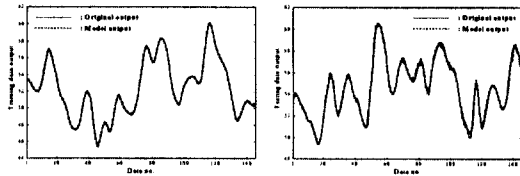


그림 4. SOPNN의 출력 데이터와 모델의 출력과의 비교

그림 5는 그림 4에서 보여준 출력 데이터와 모델의 출력과의 차이를 보여준 것이다.

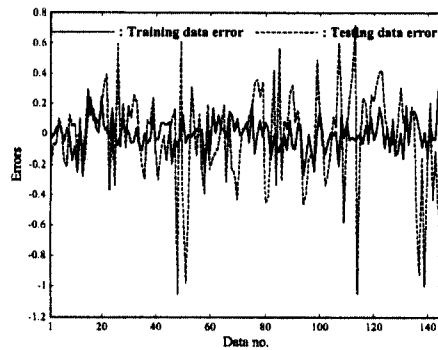


그림 5. SOPNN의 출력 데이터와 모델의 출력과의 오차

그림 6은  $\theta=0.5$ 일 때  $PI=0.010$ 이고,  $E_{PI}=0.103$ 인 경우의 모델을 나타낸 것이다. 그림에서 각 층의 노드는 다음과 같이 표시한다.

“(  $\frac{PD_A}{B \mid C}$  ) ” 여기서,  $PD_A$ 는 A번째 PD 노드를 나타낸 것이고, B는 노드로 들어오는 입력의 수를, C는 노드에 사용된 차수를 나타낸 것이다.

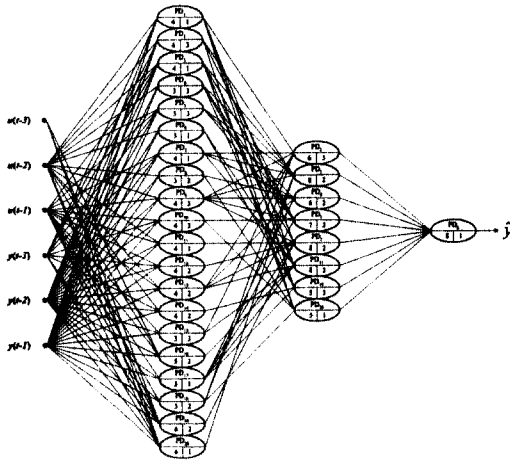


그림 6. 최적 SOPNN 구조 ( $\theta=0.5$ )

표 4는 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 수를 최적화 알고리즘인 유전자 알고리즘에 의하여 선택하여 최적의 모델을 선정하므로 모델의 정밀도가 높다. 그러므로 기존의 모델들 보다 훨씬 좋은 학습 성능결과를 얻었다.

Model	PI	EPI
Oh and Pedrycz's model[9]	0.020	0.271
Kim, et al.'s model[10]	0.034	0.244
Lin and Cunningham's model[11]	0.071	0.261

표 4. 기존 모델방법과의 동정 오차비교

## 4. 결론

본 논문에서는 데이터 수가 적거나 혹은 복잡한 비선형 요소가 많은 시스템의 체계적이고 효율적인 최적 모델을 얻기 위하여 유전자 알고리즘을 이용한 자기구성 다항식 뉴럴네트워크(SOPNN) 구조의 최적 설계방법을 제안하였다.

SOPNN 알고리즘은 선형 모델에 비해 우수한 시스템 묘사력을 가지고 있음에도 불구하고 SOPNN의 파라미터인 입력변수와 차수가 시스템의 성능과 직접적인 연관을 가지고 있음으로 객관적이고 신뢰성 높은 선택방법이 필요하다. 따라서 본 논문에서는 입력변수와 차수의 선택문제를 해결하기 위해 유전자 알고리즘으로 SOPNN의 각 노드에 인가되는 입력변수와 차수를 자동적으로 선택하였다. 그 결과 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. 또한 본 논문에서는 학습 및 테스트 데이터의 적절한 성능을 조절하여 모델의 근사화와 일반화 사이에 적절한 방향을 제시하였으며 기존의 모델보다 진화론적으로 설계된 SOPNN 모델이 더 우수한 특성과 성능이 향상됨을 알 수 있었다.

## 참고 문헌

- [1] 오성권, 김동원, 박병준, “다항식 뉴럴네트워크 구조의 최적 설계에 관한 연구”, *Trans. KIEE*, Vol. 49D, No. 3, pp. 145-156, 2000.
- [2] 오성권, 김동원, “적응다항식 뉴로-퍼지 네트워크 구조에 관한 연구”, *Trans. KIEE*, Vol. 50D, No. 9, pp. 430-438, 2001.
- [3] Dong-Won Kim, Sung-Kwun Oh and Hyun-Ki Kim, “A study on the



- Self-organizing Fuzzy Polynomial Neural Networks", *Journal of KIEE*, Vol. 11, No. 2, pp. 79-89, 2001.
- [4] Sung-Kwun Oh, Dong-Won Kim, Byoung-Jun Park, and Hyung-Soo Hwang, "Advanced Polynomial Neural Networks Architecture with New Adaptive Nodes", *Trans. on Control, Automation and Systems Engineering*, Vol. 3, No. 1, pp. 43-50, 2001.
- [5] A. G. Ivahnenko, "Polynomial theory of complex systems", *IEEE Trans. On Systems, Man and Cybernetics*, Vol. SMC-1, pp. 364-378, 1971.
- [6] S.J. Farrow, "*Self-organizing Methods in Modeling: GMDH Type Algorithms*", Marcel Dekker, New York, 1984.
- [7] J.H. Holland, "*Adaptation in Natural and Artificial Systems*", The Univesity of Michigan Press, Ann Arbor, MI., 1975.
- [8] G.E.P. Box and F.M. Jenkins, "*Time Series Analysis : Forecasting and Control*", 2nd ed. Holden-day, 1976.
- [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, issue 2, pp. 205-230, Jul. 2000.
- [10] E. Kim, H. Lee, M. Park, M. Park, "A simple identified Sugeno-type fuzzy model via double clustering", *Information Sciences*, Vol. 110, pp. 25-39, 1998.
- [11] Y. Lin, G.A. Cunningham III, "A new approach to fuzzy-neural modeling", *IEEE Trans. Fuzzy Systems*, Vol. 3, No. 2, 190-197, 1995.
- [12] 김동원, "자기구성 다항식 뉴럴네트워크의 진화론적 설계", 원광대학교 석사논문, 2002.