

GMDH 방법에 의한 FPNN 알고리즘과 폐수처리공정에의 응용

Fuzzy Polynomial Neural Network Algorithm using GMDH Method and Its Application to the Wastewater Treatment Process

오 성 권*, 황 형 수*, 안 태 천*

Sung-Kwun Oh*, Hyung-Soo Hwang*, Tae-Chon Ahn*

※이 논문은 1994년도 한국학술진흥재단의 대학부설연구소 연구과제 연구비에 의하여 연구되었음.

요 약

본 논문에서는 복잡한 비선형 시스템의 모델동정을 위해 퍼지모델링의 새로운 방법이 제안된다. 제안된 FPNN 모델링은 공정시스템의 입출력 데이터로부터 GMDH 방법과 퍼지구현규칙을 이용하여 시스템의 구조와 파라미터 동정을 구현한다. 퍼지구현규칙의 전반부 구조와 파라미터 동정을 위하여 GMDH 방법과 회귀다항식 퍼지추론 방법이 사용되고 최적 후반부 파라미터 동정을 위하여 최소자승법이 사용된다. 가스로 시계열데이터 및 하수처리시스템의 활성오니 공정 데이터가 제안한 FPNN 모델링의 성능을 평가하기 위해 사용된다. 제안된 방법이 기존의 다른 논문과 비교하여 더 높은 정확도를 가진 지능형 모델을 생성함을 보인다.

ABSTRACT

In this paper, A new design method of fuzzy modeling is presented for the model identification of nonlinear complex systems. The proposed FPNN(Fuzzy Polynomial Neural Network) modeling implements system structure and parameter identification using GMDH(Group Method of Data Handling) method and linguistic fuzzy implication rules from input and output data of processes. In order to identify premise structure and parameter of fuzzy implication rules, GMDH method and regression polynomial fuzzy reasoning method are used and the least square method is utilized for the identification of optimum consequence parameters. Time series data for gas furnace and those for wastewater treatment process are used for the purpose of evaluating the performance of the proposed FPNN modeling. The results show that the proposed method can produce the fuzzy model with higher accuracy than other works achieved previously.

*원광대학교 공과대학 제어계측공학과

I. 서 론

여러 가지 형태의 시스템을 모델링 할 때는 그 시스템에 대해 알아내기가 불가능한 변수들에 대해 알고 있는 모델설계자를 필요로 한다. 모델설계자가 이런 변수들을 추측을 통해 알아야만 한다면 좋은 예측 신뢰도를 가진 모델을 만들거란 거의 불가능하게 된다. 이런 어려움을 없애기 위해서 모델에 모델설계자의 선입관을 첨가시키지 않은 순수한 객관적 모델을 구축하게 되었다. 즉, 모델이란 단순히 데이터를 분석하여 만들어지는 것이다. 그래서 입출력 데이터의 기본이 되는 선형계의 동정, 예측문제에 관한 회귀분석의 시스템 모델링 기법 등이 유용하게 사용된다. 그러나 Modeling을 하고자 하는 시스템이 복잡하고 대규모의 구조인 경우에는 함수식으로 모델의 표현이 한정되지 않기 때문에 회귀 분석을 적용할 수 없는 경우가 있다. 일반적으로 복잡한 다변수계의 Modeling은 많은 입출력 변수 중에서 모델을 구성하는 변수를 결정하여 모델의 구조를 결정하는데 어려움이 있고 추정방법의 계수가 많으면 추정에 필요한 데이터 양이 방대해지는 문제점을 가지고 있다. 이런 문제를 해결하기 위해서 A. G. Ivakhnenko[3]가 GMDH(Group Method of Data Handling)를 제안했다. 이 방법은 입력변수의 선택 방법과 입출력 데이터의 분할방법 부분 표식의 표현 방법, 알고리즘의 종료판정과 대상 시스템에 최적화 되는 하나의 방정식을 선택한다. 최적 추정 모델을 얻기 위해서는, 수세대에 걸친 계층의 되풀이되는 계산이 필요하지만 기존의 Regression 방법보다는 적은 방정식으로 방정식의 해를 구할 수 있다. GMDH는 입출력 관계나 Model의 함수형이 특정하지 않은 비선형관계의 경우에 2변수의 2차식에 의한 부분표현식을 계층적인 조합으로 비선형 모델 추정식을 얻을 수 있다. 각 계층은 부분 표현식에 의해 회귀 분석에 적용한다, 2변수 2차식을 적용한 경우에 입출력관계의 비선형함수가 용이하게 얻어질 수 있고, 매우 적은 데이터에서 복잡하고 다변수인 비선형계의 동정, 계측이 가능하다. Hideo Tanaka 등에 의한 확신도 인자를 가진 후반부 퍼지추론규칙에 의한 방법은 기존의 간략 및 선형 퍼지추론보다 좋은 모델 성능 특성을 얻지 못한다[4][5]. 본 논문에서는 최적의 모델을 얻기 위해서 GMDH의 기본이론에 퍼

지추론식의 후반부가 Polynomial형태인 간략 퍼지추론법, Linear 추론법 및 Quadratic 추론법 등을 접목시킨 FPNN방법을 제안하였다. 제안된 방법은 기존의 방법보다 성능이 우수하다, 본 논문에서는 가스로 데이터 및 하수처리시스템의 활성오니공정을 이용하여 제안된 모델링 방법의 타당성과 모델링 결과의 정확성을 시뮬레이션을 통해 검토한다.

II. GMDH의 구조와 알고리즘

1) GMDH의 구조

모델링 하고자 하는 대상 시스템의 입력변수 x_i 와 x_j 그리고 출력 y 의 쌍에 대해서 regression 방정식을 계산함으로써 이 알고리즘을 시작한다.

$$y = A + Bx_i + Cx_j + Dx_i^2 + Cx_j^2 + Fx_i x_j$$

이 방정식을 사용하여 기존의 m 개의 입력 x_1, x_2, \dots, x_m 의 n 개의 출력 y 를 예측하는 $m(m-1)/2$ 개의 고차 변수를 얻을 수 있다. 이 입/출력 집합으로부터 regression 방정식을 찾은 후 어느 방정식이 살아남아야 하는지 찾아낸다. 이 과정에서 y 를 가장 잘 예측하는 2차 regression 모델들의 집합을 얻을 수 있다. 이제 막 계산이 끝난 각각의 이차 방정식을 이용하여 새로운 독립적 관찰결과를 얻는다. 즉 x_1, x_2, \dots, x_m 의 기존의 관찰결과를 새로이 얻은 관찰결과로 대체한다. 이 새로운 독립적 변수를 앞에 했던 것처럼 똑같은 방식으로 그 변수들을 조합시킨다. 즉, 이런 2차 regression 방정식들 모두를 계산한다. 이것을 통해서 새로운 변수로부터 y 를 예측하는 $m(m-1)/2$ 개의 regression 방정식들의 새로운 집합을 얻고, 이런 변수들은 그 이전의 방정식으로부터 얻어진 y 의 값이다. 이제 이 새로운 계산식 가운데 가장 좋은 것을 선택하고 선택된 방정식으로 새로운 독립변수를 만들어 과거의 방정식을 대체하고, 이 새로운 변수들의 쌍을 조합한다. 이 regression 방정식들이 이전의 것보다 예측 능력이 떨어지기 시작하면 이 반복 과정을 중지한다. 다시 말해서, 모델의 예측 성능이 나빠지기 시작하면 새로운 regression 방정식을 얻는 과정을 중지하게 된다. 이런 일련의 과정이 끝난 후, 가장 마지막 단계에서 얻어진 다항식들 중에서 가장 좋은 이차 다항

식을 선택한다. 우리가 얻은 다항식은 두개의 변수로 이루어진 이차 방정식의 계산 결과이다. 이 변수들은 두개 이상의 변수들의 2차식으로 구성되어 있다. 다시 말해서, 대수적 대입 법칙을 사용하면, 아래와 같은 매우 복잡한 형태의 다항식을 얻을 수 있다.

$$y = a + \sum_{i=1}^m b_i x_i + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m d_{ijk} x_i x_j x_k + \dots$$

이 식을 Ivakhnenko 다항식이라 한다.

2) 개선된 GMDH Algorithm

입력변수의 선택과 알고리즘 최종 종료판정 기준에 의해 파라미터 동정에 의한 model을 선정하게 되고, 계층적 모델추정에 의해 최종적으로 최적 model추정식을 얻는다.

공정시스템의 입출력 데이터 $(x_i, y_i) = (x_{1i}, x_{2i}, \dots, x_{mi}, y_i), i = 1, 2, \dots, N$ (데이터 총수)가 주어진 경우에, GMDH는 이 데이터의 입출력관계가 다음에 의한 종속관계 f 를 만족한다고 생각한다.

$$y = f(x_1, \dots, x_n) \tag{1}$$

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

$$\hat{y} = \hat{f}(x_1, x_2, \dots, x_n) = c_0 + \sum c_{k1} x_{k1} + \sum c_{k1k2} x_{k1} x_{k2} + \sum_{k1k2k3} c_{k1k2k3} x_{k1} x_{k2} x_{k3} + \dots \tag{2}$$

으로 표시된다. 단, c_k 는 계수이다.

GMDH는 추정 model \hat{f} 를 얻기 위해 먼저, 2변수의 부분 표현식을 구성한다. 다음에, 회귀분석으로 부분 표현식의 계수를 구하고, 중간변수 z_k 를 구한다. 중간변수에 의한 입출력 데이터를 구성시, 다음의 계층에서 같은 형태의 조작을 되풀이한다. 이 일련의 조작을 알고리즘의 종료 판정식이 만족할 때까지 되풀이한다. 알고리즘의 종료후, 얻어지는 부분표현식을 계층적으로 조합하고, 최종 추종 model \hat{f} 를 얻는다. 구체적으로, 다음 알고리즘의 절차에 의해 계산된다.

[단계 1]

출력변수 y 에 관계하는 m 개의 입력변수를 설정한다. 필요하다면 데이터를 표준화한다. 이 입력변수를 x_1, x_2, \dots, x_m 라 하자

[단계 2]

출력 y 와 m 개의 입력변수들과의 상관계수가 큰 n 개($n \leq m$)를 선택하여 model에 사용되는 입력변수로 한다.

[단계 3]

n 개의 입력변수 x 중에서 두 입력변수 x_p, x_q 를 선택하고, 다음과 같은 2변수 2차식의 부분표현식을 구성한다.

$$z_k = c_0 + c_1 x_p + c_2 x_q + c_3 x_p^2 + c_4 x_q^2 + c_5 x_p x_q \tag{3}$$

여기서, $k = 1, 2, \dots, n(n-1)/2$

여기서, c_0, c_1, \dots, c_5 는 계수이다. z_k 를 중간변수라 한다.

(3)식 대신에, 사용할 수 있는 다항식은 표 1에 나열하였다.

표 1. 1, 2, 3차 회귀다항식의 구조

Table 1. Construction of regression eqs. with degree of 1, 2, 3

No. of inputs degree	1	2	3
1	linear	bilinear	trilinear
2	quadratic	biquadratic	triquadratic
3	cubic	bicubic	tricubic

여기서, $\text{trilinear} = C_0 + C_1 x_1 + C_2 x_2 + C_3 x_3$

$$\text{triquadratic} = \text{trilinear form} + C_4 x_1 x_2 + C_5 x_1 x_3 + C_6 x_2 x_3 + C_7 x_1^2 + C_8 x_2^2 + C_9 x_3^2$$

$$\text{tricubic} = \text{triquadratic form} + C_{10} x_1 x_2 x_3 + C_{11} x_1^3 + C_{12} x_2^3 + C_{13} x_3^3$$

[단계 4]

선형 회귀분석에는, 학습 데이터를 사용하고, 계수 c_0, c_1, \dots, c_5 를 추정한다. 즉,

$$E = \sum_{i=1}^{N_i} (y_i - z_{ki})^2 \tag{4}$$

를 최소화하는 계수 c_0, c_1, \dots, c_5 를 구한다.

[단계 5]

계수 c_0, c_1, \dots, c_5 를 추정된 (3)식의 부분표현식에 평가치용 데이터를 대입하고, 평가치 데이터에 대한 (4)식의 2승오차치 E를 계산한다. 얻어진 $n(n-1)/2$ 개의 $E_1, E_2, \dots, E_{n(n-1)/2}$ 에 대해서, 값이 작은 W개를 선택하고 값이 작은순으로 나란히 교체하여 E_1, E_2, \dots, E_w 로 한다.

나머지 $n(n-1)/2 - W$ 개의 model은 버린다. 또, model 선택시 문턱값 θ_m 보다 값이 작은 model를 선택하여 값의 작은 순으로 나란히 교체하는 방법도 있다.

[단계 6]

단계 5에서 얻어진 E_i 가 다음의 부등식을 만족하는 경우에 알고리즘을 종료한다.

$$E_i \geq E_* \tag{5}$$

단, E_* 는 전층의 2승오차치 E_i 이다.

[단계 7]

(5)식을 만족하지 않는 경우, $x_{1i} = z_{1i}, x_{2i} = z_{2i}, \dots, x_{wi}, z_{wi}$ 으로서 새로운 입출력 데이터를 구성하고, 단계 3으로 간다. 이후, 단계 3부터 단계 7까지를 되풀이한다. 알고리즘이 종료할 경우, 2승오차치 E_* 를 얻고 (3)식의 부분표현식에 전층의 중간변수를 대입한다. 동일한 조작을 제1층까지 반복하고, 최종 추정 model \hat{f} 를 얻는다.

III. GMDH알고리즘에 의한 FPNN 모델구조

본 논문에서 제안한 GMDH알고리즘에 의한 FPNN 모델구조와 기존의 GMDH구조와 다른 점은 기존의 GMDH가 2변수 2차식을 사용하는 반면 제안한 GMDH알고리즘에 의한 FPNN 모델구조는 퍼지추론과 GMDH 알고리즘에 의한 PNN모델구조의 융합에 의해 생성되는 구조이다. 전건부 각 입력변수에 대한 퍼지변수의 멤버쉽함수는 가우시안 함수(Gaussian function)를 사용하고, 후건부는 단일의 상수항인 퍼지 모델과 후건부가 회귀다항식인 퍼지 모델을 사용하여 출력을 구한다.

3.1 간략 추론

후반부가 단일의 상수항만을 가지는 것으로 이와 같은 추론법을 간략추론법이라 한다. 이 모델은 식(6)의 형태를 가지는 구현규칙들로 구성되어 있다.

R^i : If x_1 is A_{i1}, \dots , and x_k is A_{ik} , then $y = a_i$

$$y^0 = \frac{\sum_{i=1}^n w_i a_i}{\sum_{i=1}^n w_i} \tag{6}$$

여기서 R^i 는 $i(i=1, \dots, n)$ 번째 규칙, $x_j(j=1, \dots, k)$ 는 입력변수, $A_{ij}(i=1, \dots)$ 는 퍼지집합의 멤버쉽 함수, a_i 는 상수이고, n 은 퍼지규칙수, 그리고 y^0 는 추론된 값이다. 후반부 파라미터 동정에서 전반부 입력변수 및 파라미터가 주어지면, PI(Performance index)를 최소화하는 최적 후반부 파라미터를 결정할 수 있다. 이 PI는 성능지수로 식 (7)로 정의된다.

$$PI = \sum_{i=1}^m \frac{(y_i - y_i^0)^2}{m} \tag{7}$$

여기서 $y_i^0 (= \sum_{i=1}^n w_i a_i / \sum_{i=1}^n w_i)$ 는 입력 $x^0 (= (x^0_1, x^0_2, \dots, x^0_k))$ 일때의 출력이며 $w^i = \prod_{j=1}^k A(x^0_j)^0$, n 은 구현 규칙수, k 는 입력변수의 수, m 은 데이터 수이다.

후반부의 파라미터는 a_i 로서 입출력 데이터가 $x_{1i}, x_{2i}, \dots, x_{ki} - y_i (i=1, 2, \dots, m)$ 의 집합으로 주어졌을 때 최소자승법에 의해 구해진다. 구조 1의 퍼지모델에 있어서 최소자승법에 의한 매개변수 추정은 식(8)와 같다.

$$\begin{aligned} V(a, m) &= 1/2 \sum_{i=1}^m \epsilon^2_i = 1/2 \sum_{i=1}^m (y_i - y^0)^2 \\ &= 1/2 \sum_{i=1}^m [y_i - \sum_{j=1}^m a_j w_{ji}]^2 = 1/2 \sum_{i=1}^m [y_i - x_i a]^2 \\ &= 1/2 \|E\|^2 \end{aligned} \tag{8}$$

$$\text{여기서 } w_{ji} = \frac{A_{j1}(x_{1i}) * \dots * A_{jk}(x_{ki})}{\sum_{j=1}^m A_{j1}(x_{1i}) * \dots * A_{jk}(x_{ki})}$$

j: 규칙수 i: 데이터수 m: 데이터 총수
n: 규칙 총수이다.

그리고, $x^T_i = [w_{ij}, \dots, w_{ni}]$, $a^T = [a_1, \dots, a_n]$,
 $Y = [y_1, y_2, \dots, y_m]^T$, $E = [\epsilon_1, \dots, \epsilon_m]^T$,
 $X = [x_1 \ x_2 \ x_m]^T$

식(8)에서 $E = Y - Xa$ 이므로 최소자승법에 의한 최소값은 식 (9)에 의해 구해진다.

$$\hat{a} = (X^T X)^{-1} X^T Y \tag{9}$$

3.2 회귀다항식 추론

후반부가 표 1과 같이 Linear, Quadratic, Cubic 다항식형태인 경우에 이 퍼지모델은 식(10)의 형태를 가지는 구현규칙들로 구성된다.

Rⁱ: If x_1 is A_{i1} , ..., and x_k is A_{ik} , then $y = f_i(x_1, \dots, x_k)$ (10)

$$f_i(x_1, \dots, x_k) = a_{i0} + a_{i1} x_1 + \dots + a_{ik} x_k$$

$$y^o = \frac{w_1 f_1(x^o_1, \dots, x^o_k) + \dots + w_n f_n(x^o_1, \dots, x^o_k)}{w_1 + w_2 + \dots + w_n}$$

여기서 Rⁱ는 i번째 규칙, x_j 는 입력변수, A_{ij} 는 퍼지집합의 멤버쉽함수, 그리고 $a_{ij}(i=1, \dots, n; j=0, \dots, k)$ 는 후반부의 파라미터 y^o 는 추론된 값이다. 후반부 파라미터는 최소자승법에 의한 최소값으로 식 (11)로 구해진다.

$$\hat{a} = (X^T X)^{-1} X^T Y \tag{11}$$

여기서, $x^T_i = [w_{ij}, \dots, w_{ni}, x_{i1} w_{i1}, \dots, x_{i1} w_{ni}, \dots, x_{ik} w_{ni}]$
 $\hat{a}^T = [a_{i0}, \dots, a_{n0}, a_{i1}, \dots, a_{n1}, \dots, a_{ik}, \dots, a_{nk}]$
 $Y = [y_1, y_2, \dots, y_m]^T$,
 $X = [x_1 \ x_2 \ x_m]^T$

IV. 시뮬레이션 및 결과 고찰

1) 가스로공정

본 절에서는 이상에서 제시된 기법의 타당성 및 유용성을 비교 평가하기 위하여 Box와 Jenkins의 가스로 시계열 입출력 데이터를 사용하여 가스로의 연소된 이

산화탄소의 농도를 모델링 한다. 모델링을 하기 위한 입력 변수는 $u(t-3)$, $u(t-2)$, $u(t-1)$, $y(t-3)$, $y(t-2)$, $y(t-1)$ 이고 출력 변수는 $y(t)$ 이다. 여기서 $u(t)$ 는 가스 흐름율이고 $y(t)$ 는 연소된 이산화탄소의 농도이다.

표 2. 각 퍼지추론방법 및 퍼지변수 개수에 의한 동정오차
Table 2. Performance index in the each fuzzy reasoning method & fuzzy variables No.

a) Simplified fuzzy reasoning method

X1 : Fuzzy variables No.

	2	3	4	5	
X2:	2	1.91419	0.33730	0.23073	0.20314
Fuzzy	3	0.27784	0.86346	0.20657	0.14307
variables	4	0.17944	0.20459	0.53734	0.14626
No.	5	0.18438	0.17082	0.13111	0.36779

b) Linear fuzzy reasoning method

X1 : Fuzzy variables No.

	2	3	4	5	
X2:	2	0.05274	0.05311	0.05081	0.04867
Fuzzy	3	0.05169	0.04280	0.04178	0.03855
variables	4	0.04742	0.04215	0.03457	0.03063
No.	5	0.04612	0.03703	0.02819	0.01857

c) Biquadratic fuzzy reasoning method

X1 : Fuzzy variables No.

	2	3	4	5	
X2:	2	0.04585	0.04385	0.03905	0.03525
Fuzzy	3	0.03953	0.03456	0.02601	0.02468
variables	4	0.03671	0.02846	0.02166	0.02408
No.	5	0.03142	0.02626	0.01991	0.01379

본 논문에서 제안한 GMDH알고리즘에 의한 FPNN 모델구조를 각 퍼지추론방식 및 각 입력변수 수에 의해 성능비교 및 멤버쉽 함수의 개수에 따른 성능비교 결과는 표 2와 같고, 기존의 퍼지모델링방법과 동정 오차비교는 표 3과 같다. 성능지수로는 $PI = \frac{1}{N} \sum_{p=1}^N$

$(y_p - \hat{y}_p)^2$ 를 사용한다.

본 논문에서 제안한 GMDH알고리즘에 의한 FPNN 모델구조에서 퍼지규칙의 전진부의 각 입력변수에 대한 퍼지변수의 멤버쉽함수는 가우시안 함수(Gaussian function)를 사용하였다.

회귀다항식 추론방식을 이용한 FPNN모델이 간략 추론 방식을 이용한 모델보다 좋은 모델링 성능을 얻었고, 퍼지변수 개수의 증가에 따라 일반적으로 좋은 성능을 얻을 수 있으나 그 시스템의 비선형성, 규칙 수 및 간단한 모델구조선택 등의 다양한 요소를 고려하여 모델구조를 선택할 수 있다.

기존의 모델보다 FPNN모델을 사용하여 모델링한 결과가 우수함을 표 3에서 보인다. 선형추론을 이용한 퍼지GMDH의 가스로 공정모델링 구조는 아래 그림 1과 같다. 각 노드는 퍼지추론 부분을 나타내고, 노드로 입력된 화살표는 두개의 입력 변수를 나타낸다. 입력변수로 $u(t)$, $u(t-1)$, $u(t-2)$, $y(t-1)$, $y(t-2)$, $y(t-3)$ 를 모두 사용하여 모델링을 하였지만 추출된 최적모델의 입력변수로 $u(t)$, $y(t-3)$, $y(t-2)$, $y(t-1)$ 만이 채택되었다.

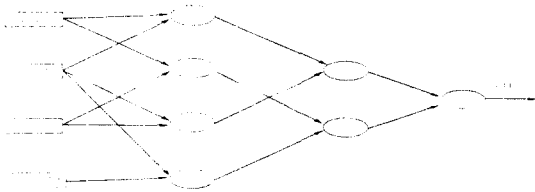


그림 1. 가스로 공정의 FPNN 모델구조(선형다항식의 경우)
Fig. 1 FPNN model structure for Gas Furnace(in the case of linear polynomial)

GMDH의 다항식을 Biquadratic방정식과 Trilinear 방정식을 사용하여 얻은 회귀다항식 퍼지추론에 의한 가스로 공정의 FPNN모델구조는 그림 2, 그림 3과 같다. Biquadratic방정식을 GMDH의 다항식으로 사용하여 얻은 모델은 그림 2와 같이 노드의 입력이 두개이며 Triquadratic방정식을 사용하여 얻은 모델의 입력은 그림 3에서와 같이 3개로 구성된다. Biquadratic방정식을 사용하여 얻은 모델이 Triquadratic방정식을 사용한 경우보다 모델의 구조가 간단함을 그림 2와 그림 3을 비교하므로써 알 수 있다.

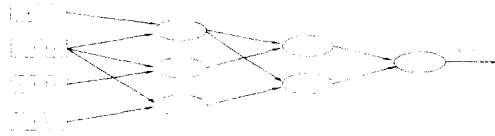


그림 2. 가스로 공정의 FPNN 모델구조(biquadratic다항식의 경우)

Fig. 2 FPNN model structure for Gas Furnace(in the case of biquadratic polynomial)

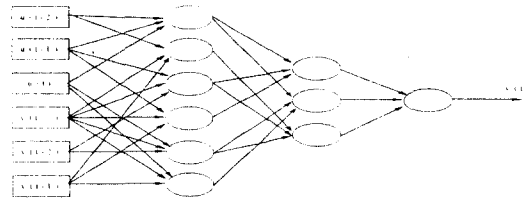


그림 3. 가스로 공정의 FPNN 모델구조(Trilinear다항식의 경우)

Fig. 3 FPNN model structure for Gas Furnace(in the case of Trilinear polynomial)

기존의 모델링 방법과 제시된 모델링 방법과의 성능비교는 표 3과 같다.

표 3. 기존의 퍼지모델링방법과 동정오차비교

Table 3. Comparison of identification error with conventional fuzzy modeling methods

모 델	평가 지수
Tong's model[11]	0.469
Pedrycz's model[12]	0.320
Xu's model[13]	0.328
Sugeno's model[14]	0.355
Oh's model[8]	0.098
Sugeno's model[15]	0.068
퍼지 GMDH[10]	0.062
FPNN(Our model)	0.039

퍼지 GMDH 동정모델의 출력과 실 데이터와의 비교는 그림 4와 같다.

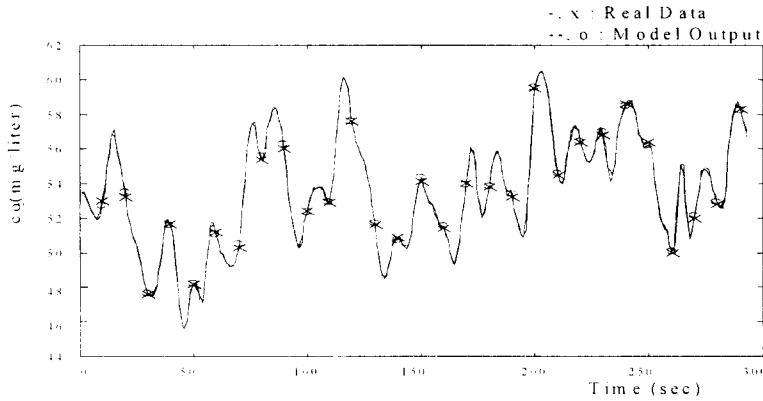


그림 4. FPNN 모델의 출력과 실 데이터의 비교
Fig. 4 Comparison of original data and output for FPNN model

2) 폐수처리공정

하폐수처리시스템에서는 활성오니공정이 일반적으로 사용되고 있다. 현재 대부분의 하폐수처리 플랜트는 제어 공정에서 조절 데이터를 얻기 위해 수학적 모델을 이용하고 있다. 그러나 이러한 수학적 모델이 정확하고 효과적으로 하폐수처리공정의 변수와 파라미터간의 관계를 설정하지 못하므로 정확한 하폐수처리공정의 모델링을 위해 본 연구에서는 수도권 하수처리장 중의 하나를 모델로 선정하여 이 처리장의 1년치 수질 데이터로부터 활성오니공정을 모델링 하였다. 활성오니공정을 모델링하기 위한 입·출력 변수는 다음과 같다. 입력은 혼합액 부유물, 잉여오니흐름율, 반응율

설정치, 용존산소 설정치이며 출력은 부유물의 농도이다. 하수처리시스템의 개요도는 그림 5와 같다.

하수처리에는 활성오니공정이 일반적으로 사용되고 있다. 활성오니공정은 침사지, 최초침전지, 폭기조 및 최종침전지를 포함한다. 침사지와 최초침전지에서는 하수 중에 포함된 현탁물질의 입자가 중력에 의해 침강된다. 폭기조에서는 수시간 연속적으로 하수 중에 공기가 흡입된다. 유기물 제거의 중심을 이루는 미생물은 폭기조의 과정에서 자연발생하는 다종류의 미생물 군이고 이를 활성오니라 한다. 응집한 미생물의 덩어리를 플록(floc)이라 한다. 활성오니는 폭기조 내 공기로부터 용해한 산소에 의해 유기물을 생물화학적으로 산화해서 증식하고 대사과정에서 유기물을 물과 탄산가스로 분해한다. 최종침전지에서 플록(floc)은 침강 제거되고 맑은 물이 3차처리 공정으로 유입된다. 한편 최종 침전지에서 침강한 미생물은 폭기조 입구에서 재순환 되어 다시 유기물 제거에 사용된다.

활성오니공정은 폭기조 및 최종침전지를 하나의 단일 모델로 선정하였다. 실제 최초침전지로 유입되는 생물학적 산소요구량(Biological Oxygen Demand : BOD) 및 유입수의 부유물(Suspended Solid : SS) 농도를 측정하고, 최종침전지에서 유출되는 생물학적 산소요구량(Effluent Biological Oxygen Demand : EBOD) 및 부유물 농도(Effluent Suspended Solid : ESS)가 주요 측정 대상이 된다. 활성오니공정의 용존산소 설정치(Dissolved Oxygen Set-Point : DOSP) 및 반응율 설정

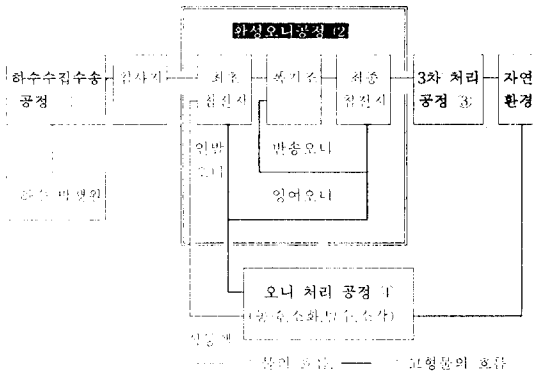


그림 5. 하수처리시스템의 개요도
Fig. 5 Configuration of sewage treatment system

치(Recycle Sludge Ratio Set-Point:RRSP)는 유입수의 생물학적 산소요구량 및 부유물 농도에 따라 유출수의 생물학적 산소요구량(EBOD), 부유물(ESS) 농도가 영향을 받으므로 조작자가 이 유출수의 농도를 소정치 이하로 낮추기 위해 사용되는 데이터이다. 유출수의 생물학적 산소요구량(EBOD) 및 부유물(ESS) 농도는 측정된 데이터의 혼합액 부유물(Mixed Liquid Suspended Solid: MLSS) 농도, 잉여오니흐름율(Waste Sludge Ratio: WSR), 반송율 설정치(RRSP), 용존산소 설정치(DOSP)에 의존하며 BOD, SS 농도는 서로 일정한 상관관계를 가지고 있다.

$$\text{평가지수는 } PI = \frac{1}{n} \sum_{i=1}^n (y - \hat{y}_p)^2 \text{이다.}$$

GMDH알고리즘에 의한 FPNN모델구조를 각 퍼지추론방식 및 퍼지변수 개수에 따른 성능비교 결과

표 4. 각 퍼지추론방법 및 퍼지변수 개수에 의한 동정오차
Table 4. Performance index in the each fuzzy reasoning method & fuzzy variables No.

a) Simplified fuzzy reasoning method

X1 : Fuzzy variables No.

	2	3	4	5
X2:	10.67944	10.22787	7.94676	4.56235
Fuzzy	9.70418	7.00801	5.86562	3.91797
variables	9.74011	6.23855	4.71324	2.06910
No.	6.21643	4.09546	2.66428	0.95624

b) Linear fuzzy reasoning method

X1 : Fuzzy variables No.

	2	3	4	5
X2:	5.58350	2.72317	2.14125	0.17251
Fuzzy	3.44119	0.57329	0.02768	0.00000
variables	1.52194	0.01681	0.00000	0.00000
No.	0.28184	0.00017	0.00000	0.00000

c) Biquadratic fuzzy reasoning method

X1 : Fuzzy variables No.

	2	3	4	5
X2:	0.83563	0.01473	0.00000	0.00000
Fuzzy	0.01434	0.00000	0.00000	0.00000
variables	0.00001	0.00000	0.00000	0.00000
No.	0.00000	0.00000	0.00000	0.00000

는 표 4와 같고, 기존의 지능형모델링방법과 동정오차비교는 표 5과 같다.

회귀다항식추론에서 후반부가 선형식인 경우 X1, X2의 각각의 퍼지변수의 개수가 3, 4개인 FPNN 동정모델의 출력과 실 데이터와의 비교는 그림 6과 같다.

간략퍼지추론 방식과 회귀다항식 퍼지추론 방식의 성능비교는 표 4에서와 같이 간략퍼지추론 방식 보다 회귀다항식 퍼지추론 방식의 성능이 우수함을 보인다. 따라서 퍼지추론방식, 규칙수, 간편한 구조및 적

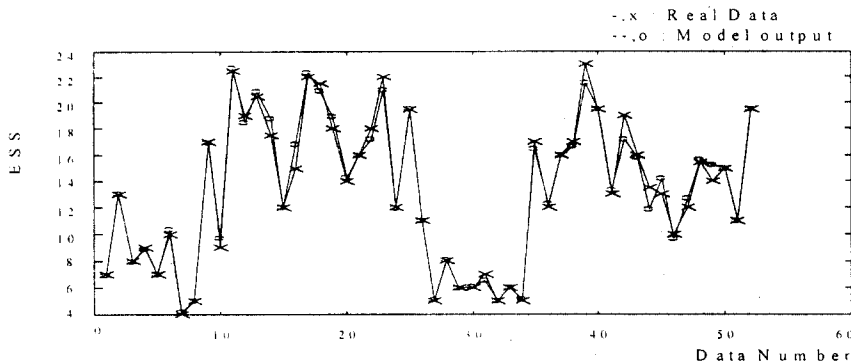


그림 6. FPNN 모델의 출력과 실 데이터의 비교
Fig. 6 Comparison of original data and output for FPNN model

절한 퍼지변수의 개수를 선택하여 모델구조를 정할 수 있다.

퍼지 GMDH를 이용하여 얻은 FPNN모델의 구조는 그림 7과 같다. 그림 7의 노드 부분은 퍼지추론 부분을 나타낸다. 노드의 입력은 전 단계 노드의 출력을 이용한다.

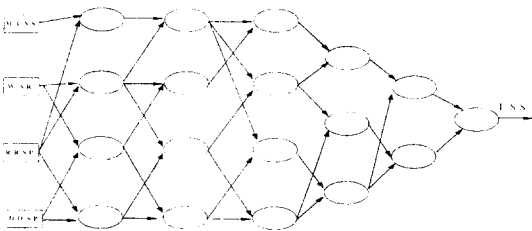


그림 7. 폐수처리 공정의 FPNN 모델구조(선형다항식의 경우)
Fig. 7 FPNN model structure for Sewage Treatment System (in the case of linear polynomial)

표 5. 기존의 지능형모델링방법과 동정오차비교
Table 5. Comparison of identification error with conventional intelligent modeling methods

모 델	평가 지수
Conventional model[8]	1.34
Fuzzy-Neural model[9]	0.56
퍼지 GMDH[10]	0.35
FPNN(Our model)	0.016

V. 결 론

본 논문에서는 데이터가 적고 비선형 요소가 많은 시스템의 수학적모델을 얻기 위해 GMDH알고리즘에 퍼지규칙의 후건부가 Polynomial형태인 간략퍼지추론, 선형퍼지추론 및 Biquadratic 퍼지추론법을 융합하여 신경회로망구조를 가진 FPNN모델구조를 제안하였다. 제안된 알고리즘들을 가스로 시계열데이터와 하수처리공정데이터를 사용하여 실공정을 모델링 함으로써 제안된 알고리즘의 성능평가를 행하였다. 제안된 FPNN알고리즘에서 퍼지규칙의 후건부 다항

식의 차수를 높이면 성능이 개선되고, 퍼지변수의 개수를 증가시키에 따라 성능이 좋아짐을 알 수 있다. 제안된 FPNN의 layer수가 어느 정도까지 증가할수록 동정오차가 감소하여 성능이 개선되지만 layer가 증가되면 FPNN 모델구조가 점점 더 복잡하게 되므로 설계 및 동정오차 설정기준에 의해 layer를 결정하게 된다. 본 논문에서 제안한 모델구조는 3층 FPNN 구조에 의한 공정시스템 모델을 구축함으로써 고부가가치의 시뮬레이터 구성에 의한 예측 및 효율적인 제어정보를 얻을 수 있으리라 기대되며, 또한 비선형 모델링을 위한 다변수 입력력 지능형 모델구조를 설계하는데도 활용될 수 있을 것이다.

참 고 문 헌

1. E. Czogala and W. Pedrycz, "On identification in fuzzy systems and its applications in control problems", *Fuzzy Sets Syst.*, Vol.6, pp.73-83, 1981.
2. T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control", *IEEE Trans. Syst. Cybern.*, Vol.SMC-15, No. 1, pp.116-132, 1985.
3. 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.
4. I. Hayashi and H. Tanaka, "The Fuzzy GMDH algorithm by possibility models and its application", *Fuzzy Sets and Systems*, 36, pp.245-258, 1990.
5. Hideo Tanaka, Katsunori and Hisao Ishibuchi, "GMDH by If-Then Rules with Certainty Factors", *Fifth IFSA World Conference*, pp.802-805, 1993.
6. 林 勳, "GMDH", *日本ファジィ學會誌* Vol.7, No. 2, pp.270-274, 1995.
7. 横出 勝則, 田中 英夫, "GMDH의 多層構造의 用いた 確信度付き의 ファジィif-then 룰", *日本ファジィ學會誌* Vol.7, No.1, pp.131-141, 1995.
8. 오 성권, 우 광방, "퍼지추론 방법에 의한 퍼지동정과 하수처리공정시스템 응용", *대한전자공학회 논문집* 제 31권 B편 제 6호, pp.43-52, 1994년 6월.
9. 오 성권, 노 석범, 박 희순, "퍼지-뉴럴네트워크에

의한 비선형공정 모델의 최적화”, 한국 퍼지 및 지능시스템 학회 춘계학술대회 논문집 제 5권 제 1호, pp.72-77, 1995년 3월.

10. 오 성권, 노 석범, 황 형수, “퍼지 GMDH 모델과 하수처리공정에서의 응용”, 한국 퍼지 및 지능시스템 학회 춘계학술대회 논문집 제 5권 제 2호, pp. 153-158, 1995년 11월.
11. R. M. Tong, “The evaluation of fuzzy models derived from experimental data”, *Fuzzy Sets Syst.*, Vol.13, pp.1-12, 1980.
12. W. Pedrycz, “An identification algorithm in fuzzy relational system”, *Fuzzy Sets Syst.*, Vol.13, pp. 153-167, 1984.
13. C. W. Xu, and Y. Zailu, “Fuzzy model identification self-learning for dynamic system”, *IEEE Trans. on Syst. Man, Cybern.*, Vol.SMC-17, No.4, pp.683-689, 1987.
14. M. Sugeno and T. Yasukawa, “Linguistic Modeling Based on Numerical Data”, *IFSA' 91 Brussels, Computer, Management & System Science*, pp. 264-267, 1991.
15. T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control”, *IEEE Trans. on Fuzzy Systems*, Vol.1, No.1, pp.7-31, 1993.



오 성 권(Sung-Kwon Oh) 정회원
 1957년 2월 13일생
 1981년: 연세대학교 전기공학과 졸업
 1983년~1989년: 금성산전연구소 (선임연구원)
 1993년: 연세대 대학원 전기공학과 졸업(공학박사)

1995년~현재: 캐나다 Manitoba대학 전기 및 컴퓨터 공학과 Post-Doc.

현재: 원광대 공대 제어계측공학과 조교수
 ※주관심분야: 시스템자동화, 퍼지논리 및 신경회로 망 응용, 모델링 및 지능형제어 등임.

황 형 수(Hyung-Soo Hwang) 정회원
 1955년 8월 2일생

1981년: 전북대학교 전기공학과 졸업
 1983년: 동 대학원 전기공학과 졸업(석사)
 1987년: 동 대학원 졸업(공학박사)
 1991년~1992년: 캐나다 Ottawa대학 전기공학과 Post-Doc.

현재: 원광대 공대 제어계측공학과 부교수
 ※주관심분야: 인공지능, 지능제어, 실시간 제어시스템 등임.



안 태 천(Tae-Chon Ahn) 정회원
 1955년 10월 11일생
 1978년: 연세대학교 전기공학과 졸업
 1980년: 동 대학원 전기공학과 졸업(석사)
 1986년: 동 대학원 졸업(공학박사)

1987년~1988년: Uppsala Univ. Post-Doc.
 1994년~1995년: Georgia Institute of Technology Visiting Prof.

현재: 원광대 공대 제어계측공학과 교수
 ※주관심분야: 적응제어, 신호처리, 지능형 모델링 및 제어 등임.