

확장된 GMDH 알고리즘에 의한 비선형 시스템의 동정

김동원*, 박병준*, 오성권*, 김현기**

* 원광대학교 전기전자공학부, ** 수원대학교 전기전자정보통신공학부

Identification of Nonlinear System using Extended GMDH algorithm

Dongwon Kim*, Byoungjun Park*, Sungkwun Oh*, Hyunki Kim**

* Division of Electrical & Electronic Engineering of Wonkwang Univ.

** Dept. of Electrical Engineering, Suwon Univ.

Abstract - The identification of nonlinear system using Extended GMDH(EGMDH) is studied in this paper. The proposed EGMDH algorithm is based on GMDH(Group Method of Data handling) method and its structure is similar to Neural Networks. The each node of EGMDH structure utilizes several types of high-order polynomial such as linear, quadratic and cubic, and is connected as various kinds of multi-variable inputs. As the operating condition changes, the parameters of EGMDH will also change, so the proposed scheme by means of the EGMDH method is capable of adapting rapidly to the changing environment. The simulation result shows that the simple nonlinear process can be modeled reasonably well by the proposed method which are simple but efficient.

식을 계산함으로써 이 알고리즘은 시작된다.

$$y = A + Bx_1 + Cx_2 + Dx_1^2 + Ex_2^2 + Fx_1x_2 \quad (1)$$

이 방정식을 사용하여 기존의 n 개의 입력 x_1, x_2, \dots, x_n 과 이에 대한 m 개의 출력 y 를 예측하는 $n(n-1)/2$ 개의 변수를 얻을 수 있다. 여기서 얻은 입출력 집합으로부터 회귀다항식을 찾는 후 어느 다항식이 살아남아야 하는지 판단 기준을 통해 결정된다. 이 과정에서 y 를 가장 잘 예측하는 2차 회귀 모델들의 집합을 얻을 수 있다. 이제 계산이 끝난 각각의 이차 다항식을 이용하여 새로운 독립적 관찰결과를 얻는다. 즉 x_1, x_2, \dots, x_n 의 기존관찰결과를 새로이 얻은 관찰결과로 대체한다. 이 새로운 독립적 변수를 앞에 했던 것처럼 똑같은 방식으로 그 변수들을 조합시킨다. 즉 이런 2차 회귀다항식들 모두를 계산한다. 이것을 통해서 새로운 변수로부터 y 를 예측하는 $n(n-1)/2$ 개의 회귀다항식들의 새로운 결과를 얻고, 이런 변수들은 그 이전의 방정식으로부터 얻은 y 값이다. 이제 새로운 계산식 가운데 가장 좋은 것을 선택하고, 선택된 다항식으로 새로운 독립변수를 만들어 과거의 다항식을 대체하고, 새로운 변수들의 쌍을 조합한다. 이 회귀다항식들이 이전의 것보다 예측능력이 떨어지기 시작하면 새로운 회귀다항식을 얻는 과정을 중지하게 된다. 이런 일련의 과정이 끝난 후, 가장 마지막 단계에서 얻어진 다항식들 중에서 가장 좋은 이차 다항식을 선택한다. 우리가 얻은 다항식은 두 개의 변수로 이루어진 이차다항식의 계산결과이다. 이 변수들은 두 개 이상의 변수들의 2차식으로 구성되어 있다. 다시 말하면, 대수적 대입 법칙을 사용하면, 아래 식 (2) 와 같은 매우 복잡한 형태의 다항식을 얻을 수 있다.

1. 서 론

근래에는 많은 연구가들이 시스템 모델링 방법을 다양한 방법으로 연구하고 있다. 그 중 기본이 되는 것이 입출력 데이터에 의한 선형계의 동정과 예측문제에 관한 회귀분석 모델링 기법과 같은 수학적 모델링이 유용하게 사용되고 있다. 그러나 모델링을 하고자 하는 시스템이 복잡하고, 대규모구조인 경우에는 함수식으로 모델의 표현이 한정되지 않기 때문에 회귀분석을 적용할 수 없는 경우가 있다. A. G. Ivakhnenko는 이런 문제점을 해결하기 위한 방법으로 2변수 2차식에 의한 부분표현식을 계층적으로 조합하여 비선형 모델 추정식을 얻을 수 있는 GMDH (Group Method of Data Handling) 알고리즘을 제안했다[1]. 그러나 자연계의 다양한 시스템을 예측하는데 있어 2변수로 국한을하여 다양한 입력변수들끼리의 상호존적인 면을 약하게 하였으며, 2차식으로는 다소 복잡한 시스템을 표현하기에는 역부족이다. 또한 간단히 표현될 수 있는 경우에도 간략화된 구조로 나타내는 것이 효과적이지만 모든 경우를 2차식으로 표현하므로 시스템에 맞는 적용력이 떨어지는 것이 단점이다. 본 논문에서는 2변수 2차식 GMDH 알고리즘의 한계를 극복한 확장된 GMDH 알고리즘을 제안한다. 다변수에 의해 변수의 개수를 확장하였으며, 2차식은 1차(linear), 2차(quadratic), 3차(cubic) 등으로 차수를 다양화하여 시스템에 맞는 구조로 적응능력을 향상시켰다. 최적의 추정모델은 수 세대를 거친 계층의 부분표현식에 의하여 얻어진다.

$$y = a + \sum_{i=1}^n b_i x_i + \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n d_{ijk} x_i x_j x_k + \dots \quad (2)$$

이 식을 Ivakhnenko 다항식이라 한다

2.2 확장된 GMDH 알고리즘

GMDH 알고리즘에 기초한 EGMDH 알고리즘은 고계 다항식과 다변수 입력에 의하여 차수 및 입력변수의 선택과 알고리즘의 최종종료판정 기준에 따라, 파라미터 동정에 의한 모델을 선정하게 되고, 계층적 모델추정에 의하여 최종적으로 최적의 모델추정식을 얻는다. EGMDH 알고리즘은 입출력 데이터 $(x_i, y_i) = (x_{1i}, x_{2i}, \dots, x_{ni}, y_i; i=1, 2, \dots, n)$ 의 관계가 다음의 종속관계 f 를 만족한다고 생각한다.

$$y = f(x_1, x_2, \dots, x_n) \quad (3)$$

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

2. 본 론

2.1 GMDH 알고리즘

모델링 하고자 하는 대상 시스템의 입력변수 x_1 와 x_2 , 그리고 출력 y 의 쌍에 대해서 식 (1)과 같은 회귀방정

$$\hat{y} = \hat{f}(x_1, x_2, \dots, x_n) \\ = c_0 + \sum C_{k1} x_{k1} + \sum C_{k1k2} x_{k1} x_{k2} + \sum C_{k1k2k3} x_{k1} x_{k2} x_{k3} + \dots \quad (4)$$

으로 표시된다. 단, C_k 는 계수이다. 연속 세대별 계층구조에 의한 EGMDH 알고리즘의 전체구조를 그림 1에 나타내었다. EGMDH에서 추정 모델 \hat{f} 를 얻기 위해, 먼저 부분표현식을 구성한다. 부분표현식으로 1차, 2차, 3차 회귀다항식을 사용한다. 다음에, 선형회귀분석으로 부분표현식의 계수를 구하고, 중간변수 z_k 를 구한다. 중간변수에 의한 입출력 데이터를 구성하여, 다음 계층에서도 이와 같은 형태의 연산을 되풀이하며, 알고리즘의 종료 판정식을 만족할 때까지 계속 반복한다. 알고리즘을 종료한 후 얻어지는 부분표현식은 계층적으로 조합하고, 최종 추정 모델 \hat{f} 를 얻는다. EGMDH는 각 세대의 부분표현식으로서의 다항식을 구성하는 방법에 따라 다음과 같은 각각의 경우로 이루어진다.

Case 1: 서로 다른 차수 혹은 변형된 형태의 차수를 가지고, 동일 입력 변수 수를 각 세대의 부분표현식으로 설정하는 구조로 1세대에서 2변수 1차식으로 입력 변수 수와 차수를 정한다면, 다음 세대에서는 변수의 수를 2변수로 동일하게 설정하고, 차수를 변화시킴으로써 계층끼리의 상호 연계성을 고려한 구조이다.

Case 2: 각 세대의 부분표현식으로 동일 차수와 서로 다른 입력변수의 수를 가지는 구조로 1세대에서의 부분표현식과 2세대 이상에서의 부분표현식이 서로 다른 구조를 띄고 있는 경우로 1세대를 1변수 2차식으로 설정했다면, 2세대 이상에서는 2변수 2차식 구조를 형성하며 복잡한 고계 회귀다항식까지 사용할 수 있다.

Case 3: 서로 다른 차수 혹은 변형된 형태의 차수를 가지고, 서로 다른 입력변수 수로 각 세대의 부분표현식의 구조를 설정하는 경우로 1세대에서의 부분표현식과 2세대이상의 부분표현식의 구조가 서로 다른 형태를 띄고 있을 뿐만 아니라, 입력 변수 수와 부분표현식의 차수를 서로 다르게 설정할 수 있는 구조이다.

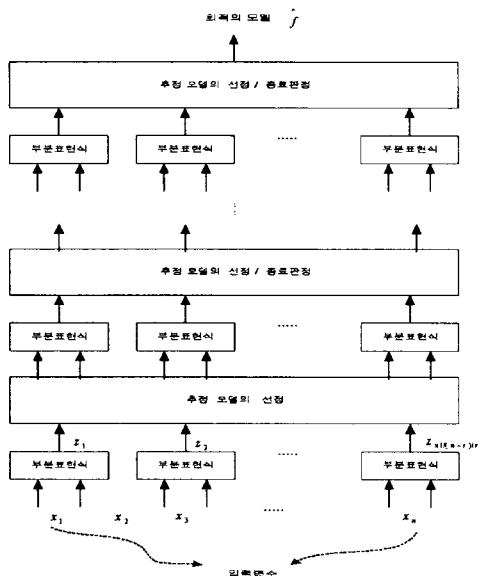


그림 1. 확장된 GMDH모델의 전체구조
Fig. 1. Overall scheme of Extended GMDH structure

2.3 시뮬레이션 및 결과고찰

본 논문에서 제안한 EGMDH 알고리즘에 의한 단계별 설계 절차를 이용하여 Sugeno가 사용한 데이터[2]를 이용한다. 모델링하기 위한 입력변수로는 x_1, x_2, x_3, x_4 이고 출력변수는 y 이다. 여기서 x_4 는 dummy 변수로 입력 변수에 부가되어 있다. 총 40개의 데이터중 처음 20개의 데이터는 학습에 이용하고 나머지 20개의 데이터는 테스트에 이용한다. EGMDH 알고리즘에 의한 모델의 성능지수로는 식 (5)에 나타난 것을 사용한다. 실제 시스템의 출력 데이터와 모델링 데이터간의 차이를 의미하는 $Error(E)$ 로 모델의 정밀도를 평가할 수 있다.

$$E = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} * 100(\%) \quad (n: \text{데이터 수}) \quad (5)$$

그림 2에서는 EGMDH 모델의 부분표현식 구조를 세대로 구분하여 변화시킨 것의 결과이다. 즉 부분표현식으로써의 다항식이 서로 다른 차수 및 변형된 차수를 가지고, 동일 입력변수를 가지는 EGMDH 구조의 결과를 나타낸 것으로 2변수일 때 Type1(1차식)에서 Type2(2차식)로 변화되는 구조의 출력값을 나타내었다.

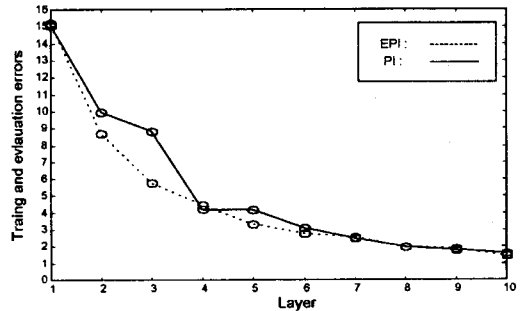


그림 2 학습 및 평가 성능지수

Fig. 2. Performance index for training and evaluation errors

그림 3은 1세대와 2세대 이상의 입력변수가 2변수에서 3변수로 변환할 때 Type 2의 다항식구조를 가진 EGMDH의 결과값을 나타내었다.

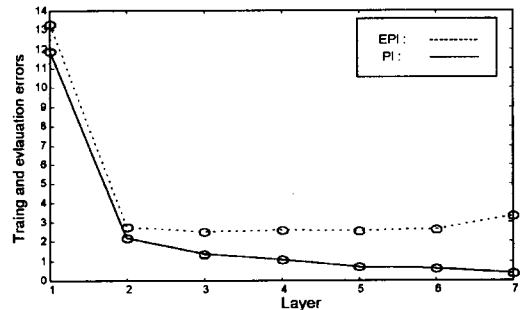


그림 3 학습 및 평가 성능지수

Fig. 3. Performance index for training and evaluation errors

1세대와 2세대이상의 입력변수가 2변수에서 3변수로 변환할 때 EGMDH 다항식의 구조가 Type 1에서 Type 2로 변환할 때의 결과값이 그림 4에서 보여진다.

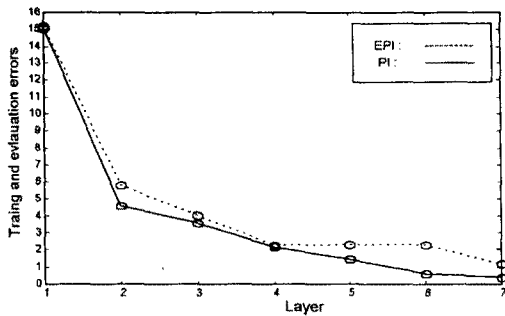


그림 4 학습 및 평가 성능지수

Fig. 4. Performance index for training and evaluation errors

제안된 모델의 결과값과 기존 모델방법의 동정에러를 표 1에 나타내어 성능지수를 비교하였다.

표 1. 기존 모델방법과 동정 에러 비교

Table 1. Comparison of identification error with previous modeling method

Model	PI(%)	EPI(%)
선형모델 [4]	12.7	11.1
GMDH [4]	4.7	5.7
Sugeno model Fuzzy I [4]	1.5	2.1
Sugeno model Fuzzy II [4]	1.1	3.6
Our model	Case 1	1.589
	Case 2	0.685
	Case 3	0.396

Case 1은 EGMDH가 입력변수로 2입력을 가지며 1세대에서 Type 1, 2세대 이상에서 Type 2의 구조로 이루어진 경우이며, Case 2는 1세대에서 2입력을 가지며, 2세대 이상에서 3입력으로 Type 2의 구조로 이루어진 경우이다. Case 3은 1층과 2층 이상에서 입력변수로 3입력과 2입력을 가지며, Type 1에서 Type 2로 변하는 구조이다.

3. 결 론

본 논문에서는 데이터가 비교적 적고 복잡한 비선형 시스템의 체계적이고 효과적인 모델링을 위하여 기존의 GMDH와 이를 확장, 개선한 GMDH(EGMDH) 알고리즘을 제안하였다. EGMDH 모델의 다층구조에 의한 모델의 표현능력과 함께 다변수의 비선형시스템에 대하여 동정과 예측이 가능하고, 근사화와 일반화를 고려한 모델 구축을 통하여 근사 및 예측성능의 향상을 꾀하였다.

감사의 글

이 논문은 1999년 한국학술진흥재단의 학술연구비에 의하여 지원되었음

[참 고 문 헌]

[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] G. T. Gang and M. Sugeno, "Fuzzy modeling", *Trans. Soc. Instrum. Control Eng.* 23(6), pp. 650-652, 1987(in Japanese)

[3] T. Takaki and M. Sugeno, "Fuzzy Identification of Systems and Its Application to Modeling and Control", *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, VOL. SMC-15, NO.1, pp.116-132, 1985.

[4] 菅野道夫(譯:朴致用·崔恒植), "퍼지제어 시스템", pp. 156, 1990

[5] 林 勳, "GMDH", *日本 ファジィ 學會誌* Vol.7, No.2, pp.270-274, 1995.

[6] Rok Soo Sang, Sung Kwun Oh, Tae Chon Ahn and Kul Hur, "A Fuzzy Model on the PNN Structure", *'98 The Third Asian Fuzzy Systems Symposium(AFSS)*, pp.83-86, June, 18-21, 1998.

[7] Ho Sung Park, Sung Kwun Oh, Tae Chon Ahn and Witold Pedrycz, "A Study on Multi-layer Fuzzy Polynomial Inference System based on an Extended GMDH algorithm", *'99 IEEE International Fuzzy System Conference Proceedings*, Vol. 1, pp. 1354-359.

[8] Byoung Jun Park, Sung Kwun Oh and Witold Pedrycz, "Fuzzy Identification by means of Partitions of Fuzzy Input Space and an Aggregate Objective Function", *'99 IEEE International Fuzzy System Conference Proceedings*, Vol. 1, pp. 1480 - 485.

[9] 오성권, "퍼지모델 및 제어이론과 프로그램", 기다리출판사, 1999년 3월.