

수정된 GMDH 알고리즘 기반 다층 퍼지 추론 시스템에 관한 연구

박병준, 박춘성, 오성권
원광대학교 전기전자공학부, 전라북도 익산시 신룡동 344-2 ☎ 570-749

A Study on Multi-layer Fuzzy Inference System based on a Modified GMDH Algorithm

Byoungjun Park, Chunseong Park, and Sungkwan Oh
Division of Electrical and Electronic Engineering, Wonkwang Univ., Iksan, KOREA

Abstract - In this paper, we propose the fuzzy inference algorithm with multi-layer structure. MFIS(Multi-layer Fuzzy Inference System) uses PNN(Polynomial Neural networks) structure and the fuzzy inference method. The PNN is the extended structure of the GMDH(Group Method of Data Handling), and uses several types of polynomials such as linear, quadratic and cubic, as well as the biquadratic polynomial used in the GMDH. In the fuzzy inference method, the simplified and regression polynomial inference methods are used. Here, the regression polynomial inference is based on consequence of fuzzy rules with the polynomial equations such as linear, quadratic and cubic equation. Each node of the MFIS is defined as fuzzy rules and its structure is a kind of neuro-fuzzy structure.

We use the training and testing data set to obtain a balance between the approximation and the generalization of process model. Several numerical examples are used to evaluate the performance of the our proposed model.

1. 서 론

일반적으로 복잡한 다변수계의 Modeling은 많은 입력 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[1]가 GMDH(Group Method of Data Handling)를 제안했다. 이 방법은 입력변수의 선택 방법과 입력력 데이터의 분할방법 부분 표식의 표현 방법, 알고리즘의 종료판정과 대상 시스템에 최적화 되는 하나의 방정식을 선택한다. 최적 추정 모델을 얻기 위해서는, 수세대에 걸친 계층의 되풀이 되는 계산이 필요하지만 기존의 Regression 방법보다는 적은 방정식으로 방정식의 해를 구할수 있다.

퍼지 추론 방법을 이용한 모델링과 예측은 동정 방법에 따라 많은 연구가 수행되어 왔다. 언어적 접근방식에서 Tong은 논리적 조사 방법에 의해 가스로 공정을 동정하였고[4], B. Li등은 이 방법의 수정으로 더 좋은 결과를 얻는 방법[5]과 결정 테이블에 기초한 자기 학습 알고리즘을 제안하였다. 퍼지관계 방정식에 기초한 접근방식에서, Sugeno는 자신이 제안한 선형 퍼지추론식에 경험적 방법을 통한 구조동정으로 커플렉스법과 최소자승법에 의해 시스템을 동정하였으나[6]. 전반부 구조동정에서 반복적인 퍼지분할에 의한 경향에 의존하는 문제점이 있다. 이를 모든 동정알고리즘은 주어진 공정의 학습 데이터에 기초함으로써 보다 좋은 성능 즉 근사화 능력에 초점이 맞추어졌다.

본 논문에서는 최적의 모델을 얻기 위해서 GMDH의 기본이론에 퍼지추론식의 후반부가 Polynomial형태인

간략 퍼지추론법, Linear 추론법 및 Quadratic 추론법 등을 접목시킨 MFIS(Multi-layer Fuzzy Inference System)를 제안한다. 또한, 기존의 학습 데이터뿐만 아니라 테스트 데이터를 고려한 모델의 성능, 즉, 근사화 능력과 예측능력 모두를 고려하여 실 공정에의 적용의 유용성에 그 방향을 제시하였다. Mackey-Glass 시계열의 데이터를 이용하여 제안된 모델의 예측성능을 판단하고, Wang[7]의 새로운 학습 알고리즘에 의한 시계열 예측과 Jang과 Sun[8]의 ANFIS(Adaptive Network-based Fuzzy Inference System) 모델에 의한 예측성능과 비교한다.

2. 본 론

2.1 개선된 GMDH 알고리즘

2.1.1 GMDH의 구조

$$y = A + Bx_i + Cx_j + Dx_i^2 + Ex_j^2 + Fx_ix_j, \quad (1)$$

위의 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 고차변수를 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. 이 새로운 독립적 변수를 앞에 했던 것처럼 똑같은 방식으로 그 변수들을 조합시킨다. 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. 새로운 계산식 가운데 가장 좋은 것을 선택하고 선택된 방정식으로 새로운 독립변수를 만들어 과거의 방정식을 대체하고, 이 새로운 변수들의 쌍을 조합한다. regression 방정식들이 이전의 것보다 예측성능이 떨어지기 시작하면 이 반복 과정을 중지한다. 가장 마지막 단계에서 얻어진 다항식들 중에서 가장 좋은 이차 다항식을 선택하면, 아래와 같은 매우 복잡한 형태의 다항식을 얻을 수 있다.

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

이 식을 Ivakhnenko 다항식이라 한다.

2.1.1 개선된 GMDH 알고리즘

[단계 1] 출력변수 y 에 관계하는 m 개의 입력변수를 설정한다. 필요하면 데이터를 규준화한다. 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_i, x_j 를 선택하고, 다음과 같은 2변수 2차식의 부분표현식

을 구성한다.

$$z_k = c_0 + c_1x_p + c_2x_p + c_3x_p^2 + c_4x_p^3 + c_5x_p x_q \quad (3)$$

여기서, $k = 1, 2, \dots, n(n-1)/2$ 이고 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 회귀 다항식의 구조

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

$$\text{trilinear} = C_0 + C_1x_1 + C_2x_2 + C_3x_3$$

$$\text{triangular} = \text{trilinear form}$$

$$+ C_4x_1x_2 + C_5x_1x_3 + C_6x_2x_3 + C_7x_1^2 + C_8x_2^2 + C_9x_3^2$$

$$\text{tricubic} = \text{triangular form}$$

$$+ C_{10}x_1x_2x_3 + C_{11}x_1^3 + C_{12}x_2^3 + C_{13}x_3^3$$

[단계 4] 선형 회귀분석에는 학습 데이터를 사용하고, 식(4)를 최소화하는 계수 c_0, c_1, \dots, c_5 를 구한다.

$$E = \sum_{i=1}^{N_i} (y_i - z_k)^2 \quad (4)$$

[단계 5] 계수 c_0, c_1, \dots, c_5 를 추정한 (3)식의 부분표현식에 평가용 데이터를 대입하고, 평가 데이터에 대한 (4)식의 2승오차치 E 를 계산한다. 얻어진 $n(n-1)/2$ 개의 $E_1, E_2, \dots, E_{n(n-1)/2}$ 에 대해서, 값이 작은 W 개를 선택하고 값이 작은순으로 나란히 교체하여 E_1, E_2, \dots, E_n 로 한다. 나머지 $n(n-1)/2 - W$ 개의 모델은 버린다.

[단계 6] 단계 5에서 얻어진 E_1 이 다음의 부등식을 만족하는 경우에 종료한다.

$$E_1 \leq E^* \quad (5)$$

단, E^* 는 전총의 2승 오차치 E_1 이다.

[단계 7] (5)식을 만족하지 않는 경우, $x_{1i} = z_{1i}, x_{2i} = z_{2i}, \dots, x_{ni} = z_{ni}$ 으로서 새로운 입력 출력 데이터를 구성하고, 단계 3으로 간다. 이후, 단계 3부터 단계 7까지를 되풀이한다. 알고리즘이 종료할 경우, 2승오차치 E^* 를 얻고 (3)식의 부분표현식에 전총의 중간변수를 대입한다. 동일한 조작을 제1층까지 반복하고, 최종 추정 모델 j 를 얻는다.

2.2 MFIS 모델의 구조와 알고리즘

본 논문에서 제안한 GMDH 알고리즘에 의한 MFIS 모델구조와 기존의 GMDH구조와 다른점은 기존의 GMDH가 2변수 2차식을 사용하는 반면 제안한 MFIS 모델구조는 퍼지추론과 GMDH 알고리즘에 의한 PNN 모델구조의 융합에 의해 생성되는 구조이다. 전반부 각 입력변수에 대한 퍼지변수의 멤버쉽함수는 가우시안 함수(Gaussian function)를 사용하고, 후반부는 단일의 상수항인 퍼지 모델과 후반부가 회귀다항식인 퍼지 모델을 사용하여 출력을 구한다.

2.2.1 간략 추론

후반부가 단일 상수항만을 가지는 것으로, 이 퍼지모델은 식(6)과 같은 형태를 가지는 구현규칙들로 구성된다. 추론된 값 y^* 은 식(7)와 같다.

$$R^n \text{ IF } x_1 \text{ is } A_{n1} \text{ and } \dots \text{ and } x_k \text{ is } A_{nk} \text{ Then } a_{n0} \quad (6)$$

$$y^* = \frac{\sum_{j=1}^n w_{nj} y_j}{\sum_{j=1}^n w_{nj}} = \frac{\sum_{j=1}^n w_{nj} a_{nj}}{\sum_{j=1}^n w_{nj}} \quad (7)$$

후반부의 파라미터는 a_{nj} 로써, 최소자승법에 의해 구해진다. 최소자승법에 의한 후반부 파라미터의 동정은 식(8)에 의해 구해진다.

$$\hat{a} = (X^T X)^{-1} X^T Y \quad (8)$$

2.2.1 회귀다항식 추론

후반부가 표 1과 같이 Linear, Quadratic, Cubic 다항식형태인 경우에 이 퍼지모델은 식(9)의 형태를 가지는 구현규칙들로 구성된다.

$$R^n \text{ IF } x_1 \text{ is } A_{n1} \text{ and } \dots \text{ and } x_k \text{ is } A_{nk} \text{ Then } y = f_n(x_1, \dots, x_k) \quad (9)$$

추론된 값 y^* 은 다음과 같다.

$$y^* = \frac{\sum_{j=1}^n w_{nj} y_j}{\sum_{j=1}^n w_{nj}} = \frac{\sum_{j=1}^n w_{nj} (a_{nj} + a_{nj}x_1 + \dots + a_{nj}x_k)}{\sum_{j=1}^n w_{nj}} \quad (10)$$

최소자승법에 의한 후반부 파라미터의 동정은 구조1과 같이 식(8)에 의해 구해진다.

전체 MFIS의 구조는 그림 1과 같다. 각 노드에는 퍼지 추론 시스템이 내재되어 있고, 퍼지 추론 시스템에서는 구조 1부터 4까지의 회귀 다항식의 후반부 구조를 가진다.

$$\text{구조 1} = a_{nj}$$

$$\text{구조 2} = a_{nj} + a_{nj}x_1 + a_{nj}x_2$$

$$\text{구조 3} = a_{nj} + a_{nj}x_1 + a_{nj}x_2 + a_{nj}x_1x_2$$

$$\text{구조 4} = a_{nj} + a_{nj}x_1 + a_{nj}x_2 + a_{nj}x_1^2 + a_{nj}x_2^2 + a_{nj}x_1x_2$$

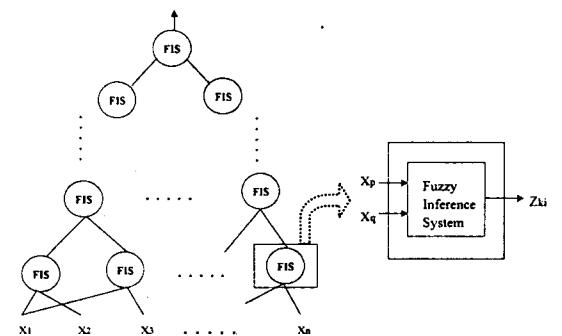


그림 1. MFIS의 전체 구조

2.3 Mackey-Glass Time Series

Mackey-Glass time series는 예측알고리즘의 성능 평가를 위한 대표적인 예로서 사용되는 시계열로 다음 식과 같은 비선형 동력학 방정식에 의해 생성된다[9].

$$\dot{x}(t) = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \quad (11)$$

기존의 방법들과 비교를 위해 논문에서는 미래값 $x(t+6)$ 의 예측성능을 알펴본다. 위의 방정식 (11)에 대한 해를 찾기 위한 방법으로 4차 Runge-Kutta 방법을 적용한다. time step은 0.1, 초기조건 $x(0)$ 은 1.2로, 시간지연상수 τ 는 17로 놓고 $x(t)$ 를 발생시킨다. 제안된 모델의 학습을 위해 $x(118)$ 부터 $x(617)$ 까지 500개의 시계열 데이터를 사용하였고, 예측성능의 평가를 위해 $x(618)$ 부터 $x(1117)$ 까지 500개의 시계열 데이터를 예측치와 비교하였다. 시뮬레이션의 예측성능을 위해, 잘 알려진 성능지수 RMSE(Root Mean Squared Error)를 이용한다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x(i) - \hat{x}(i))^2} \quad (12)$$

본 논문에서는 제안한 GMDH 알고리즘에 의한 MFIS 모델구조로서 퍼지 규칙의 전반부의 각 입력변수에 대한 멤버쉽함수는 가우시안 함수를 사용한다. Mackey-Glass 시계열에 대하여 MFIS 모델 구조를 각 퍼지 추론 방식 및 멤버쉽함수의 개수에 따른 결과는 표 2와 같다.

표 2. MFIS 모델의 동정 오차 ($x(t+6)$)

Structure	Fuzzy Variables No.	Training Error	Testing Error
Structure 1	2 × 2	0.0918	0.0937
	2 × 3	0.0557	0.0574
	3 × 2	0.0436	0.0458
	3 × 3	0.0374	0.0387
Structure 2	2 × 2	0.0140	0.0158
	2 × 3	0.0139	0.0144
	3 × 2	0.0126	0.0134
	3 × 3	0.0114	0.0158
Structure 3	2 × 2	0.0130	0.0146
	2 × 3	0.0125	0.0164
	3 × 2	0.0116	0.0136
	3 × 3	0.0143	0.0157
Structure 4	2 × 2	0.0251	0.0253
	2 × 3	0.0240	0.0244
	3 × 2	0.0235	0.0244
	3 × 3	0.0218	0.0233

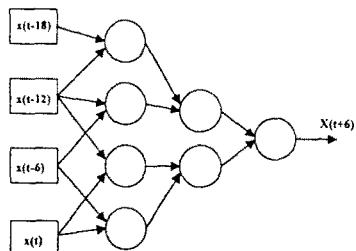


그림 2. 예측을 위한 MFIS 구조 (구조 2의 경우)

GMDH의 다항식에 구조 2를 사용하여 얻은 회귀다항식 퍼지 추론에 의한 Mackey-Glass 시계열의 MFIS 모델 구조는 그림 2과 같다. Mackey-Glass 시계열의 미래의 여섯 번째의 값을 예측한 결과로서, 그림 3은 선형 다항식 구조를 가지는 MFIS을 적용한 결과를 살펴볼 수 있으며, 그림 4는 평가용 데이터에 대한 예측 결과를 보여준다.

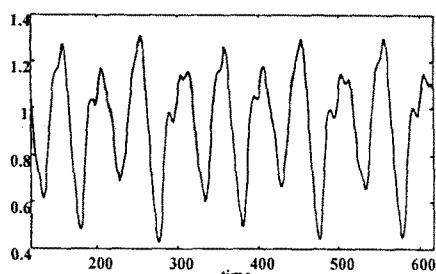


그림 3. Mackey-Glass 시계열에 대한 예측결과 ($x(t+6)$) (실선-실제치, 점선-예측치)

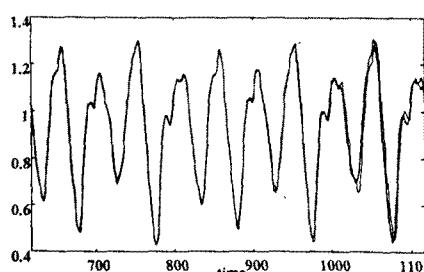


그림 4. 평가용 데이터에 대한 예측결과 ($x(t+6)$) (실선-실제치, 점선-예측치)

표 3. 기존의 퍼지 모델링 방법과 예측오차 비교

Model	Rules No.	RMSE
Wang[7]	7	0.0447
	11	0.0280
	15	0.0194
	19	0.0160
	23	0.0137
	27	0.0117
	31	0.0107
ANFIS[8]	16	0.007
Our Model	3	0.0130

Wang[7]이 제안한 새로운 알고리즘에 의한 예측성능 및 Jang과 Sun이 제안한 ANFIS[8] 모델의 예측성능과 논문에서 제안한 시계열 예측능력 비교는 표 3과 같다. Wang의 방법보다는 우수함을 알 수 있고 ANFIS 모델보다는 다소 능력이 떨어짐을 보여주지만, 많은 퍼지 규칙을 사용하였기 때문에 실제 이용하기에 구현의 복잡성이 따른다. 하지만 제안된 GMDH의 회귀다항식을 가지는 MFIS는 퍼지 구현 규칙이 단순하여 우수한 비선형 시계열 예측능력을 발휘한다.

3. 결 론

본 논문에서는 데이터가 적고 비선형 요소가 많은 시스템의 수학적모델을 얻기 위해 GMDH알고리즘에 퍼지 규칙의 후반부가 Polynomial형태인 간략퍼지추론, 선형퍼지추론 및 Biquadratic 퍼지추론법을 융합하여 신경회로망구조를 가진 MFIS 모델구조를 제안하였다. 제안된 알고리즘들을 Mackey-Glass 시계열에 대한 예측 성능을 측정하여 모델의 성능을 평가하였다. 시뮬레이션에서 보여졌듯이 기존의 방법보다 퍼지 구현 규칙이 간단하면서 우수한 예측성능을 발휘하고 있다.

(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] I. Hayashi and H. Tanaka, "The Fuzzy GMDH algorithm by possibility models and its application". *Fuzzy Sets and Systems*, 36, pp.245-258, 1990.
- [3] Hideo Tanaka, Katsunori and Hisao Ishibuchi, "GMDH by If-Then Rules with Certainty Factors". *Fifth IFSA World Conference*, pp.802-805, 1993.
- [4] R. M. Tong, "The evaluation of fuzzy models derived from experimental data". *Fuzzy Sets Syst.*, Vol.13, pp.1-12, 1980.
- [5] C.W. Xu, "Fuzzy system identification". *IEEE Proceeding* Vol. 126, No. 4, pp.146-150, 1989.
- [6] 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.
- [7] L.X. Wang and J.M. Mendel, "Generating fuzzy rules from numerical data, with applications," *IEEE Trans. on Systems, Man, and Cybern.*, 22 no. 6, pp. 1414-1427, 1992.
- [8] J.R. Jang and C. Sun, "Predicting Chaotic Time Series with Fuzzy If-Then Rules," *2nd IEEE Inter. Conf. Fuzzy Systems*, San Francisco, pp. 1079-1084, 1993.
- [9] M.C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems. *Science*, 197:287-289, July 1977.