

---

# 피드백에 의한 GMDH 알고리즘 성능 향상에 관한 연구

홍연찬\*

A Study on the Performance Improvement of GMDH Algorithm by Feedback

Yeon Chan Hong\*

---

본 논문은 2008년도 인천대학교 학술연구구조성비 지원에 의하여 수행되었음

---

## 요 약

복잡한 비선형 시스템을 예측하기 위하여 GMDH(Group Method of Data Handling) 알고리즘을 사용할 수 있다. 기존의 GMDH 알고리즘은 정해진 절차에 의해 입력층부터 중간층들을 거쳐 출력층에서 시스템의 예측 출력을 생성한다. 각 층의 출력은 전 층의 출력에 의해서만 생성된다. 그러나 전형적인 GMDH 알고리즘에서 층별로 최적의 구조가 결정되지만 전체적으로는 최적의 구조가 결정되지 않을 수도 있다는 문제점을 해결하기 위해 예측된 출력을 실제의 출력과 비교하여 그 에러를 피드백하여 전체적으로 최적의 구조를 가지는 GMDH 예측 모델을 구성함으로써 보다 정확한 예측이 가능하도록 하였다. 제안된 알고리즘이 기존의 알고리즘보다 성능이 향상된 것을 컴퓨터 시뮬레이션을 통해 검증하였다.

## ABSTRACT

The GMDH(Group Method of Data Handling) algorithm can be used to predict the complex nonlinear systems. The traditional GMDH algorithm produces the predicted output of the system model in the output layer through the input layer and the intermediate layers as the prescribed process. The outputs of each layer are produced only by the outputs of the former layer. However, in the traditional GMDH algorithm, though the optimal structure of each layer is derived, the overall structure may not be derived optimally. To overcome this problem, GMDH prediction model which has the overall optimal structure is constructed by feeding back the error between the predicted output and the real output. This can make the prediction more precise. The capability improvement of the proposed algorithm compared to the traditional algorithm is verified through computer simulation.

## 키워드

GMDH 알고리즘, 비선형 시스템, 예측 출력, 피드백, 컴퓨터 시뮬레이션

## Key word

GMDH algorithm, nonlinear system, predicted output, feedback, computer simulation

## I. 서 론

GMDH 알고리즘은 경제, 생태, 의학 진단, 신호 처리, 제어 시스템 등의 분야에서 불확실성과 비선형성을 처리하기 위해 널리 사용되고 있다[1,2,3]. 2방향 순환 GMDH[4]와 수정된 GMDH[5] 등은 예측을 위해 동적 시스템을 성공적으로 모델링할 수 있었다.

경제지수 또는 전력수요 예측을 위한 시계열 함수는 일반적으로 여러 가지 요인에 의해 불규칙하게 변동되고, 노이즈가 많은 특성이 있어 정확하게 예측하는 데 어려움이 있다. 현재까지 대부분의 시계열 함수 예측은 신경회로망, 퍼지이론 등과 같은 인공지능 기법을 적용한 방법이 활발히 연구되고 있다[6,7]. 이러한 예측 기법은 예측을 위한 데이터를 모두 처리하므로 많은 계산량으로 인해 장기적인 예측에는 계산 시간이 많이 소요된다는 단점이 있다. 그러나 각 층(layer)에서 유용한 데이터를 자체적으로 추출하는 한계값을 갖는 다층의 알고리즘을 통하여 정확성을 증가시키기 위해 해로운 정보를 제거하는 GMDH 알고리즘은 데이터의 효율적인 활용과 정확성에서 우수한 특성을 보이며 이 알고리즘을 적용한 연구가 활발히 진행되고 있다[8,9].

그러나 GMDH 알고리즘은 자기 구성(self organization) 기능에 의해 층 별로 구조가 결정되어 전체 구조를 결정하는 방식을 사용하므로 층 별로는 원하는 모델링 특성을 가질 수 있지만 전체적으로는 최적의 모델링 특성을 나타내지 못할 수도 있다. 따라서 본 논문에서는 기존의 수정된 GMDH 알고리즘에 피드백 기능을 추가하고 신경망에서 사용하는 LMS(least mean square) 학습 법칙을 적용하여 노드의 연결 강도를 조정함으로써 모델링 성능을 향상시켜서 더 우수한 예측 성능을 얻고자 한다. 제안된 알고리즘이 기존의 알고리즘보다 성능이 향상된 것을 전력 수요 예측을 통해 검증하였다.

## II. 피드백에 의한 GMDH 알고리즘

GMDH 알고리즘은 다단계(multi-layer)적 구조를 가지면 각 단계의 입력 데이터로부터 비선형 변형을 하

는데, 그림 1과 같이 두 개의 선정된 변수의 2차 다항식을 이룬다.

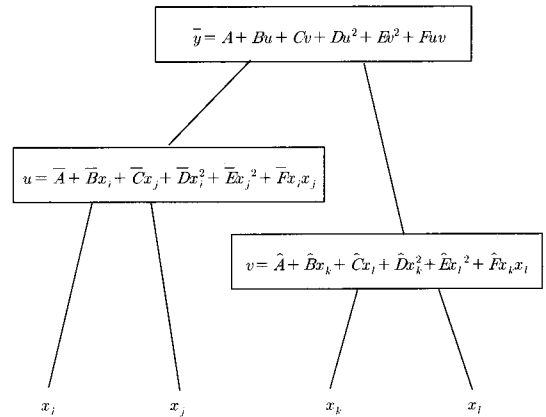


그림 1. 변수 전파의 구조.  
Fig. 1. Structure of variable propagation.

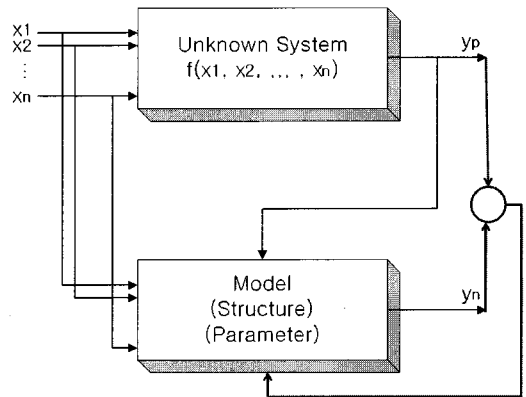


그림 2. GMDH 알고리즘을 이용한 근사 모델.  
Fig. 2. Approximate model using GMDH algorithm.

GMDH 알고리즘은 다변량 회귀 모델로 대표되는 통계적 모델을 대상으로 하는 그림 2와 같은 시스템의 입출력 정보를 기초로 하여 근사 모델을 만드는 것이다. 본 논문에서는 그림 2에 나타내었듯이 구성된 모델의 출력과 실제의 출력을 비교하여 그 피드백된 값을 이용해 LMS 학습 법칙을 적용하여 GMDH 알고리즘에서 노드의 연결 강도 역할을 하는 파라미터들을 제조정함으로써 예측 성능을 개선하고자 한다.

일반적으로 비선형 함수의 입출력 관계는 (1)과 같이 Volterra 급수로서 표현될 수 있다.

$$\begin{aligned}
 y_t = & \int_0^t h_1 x(t-\tau) d\tau \\
 & + \int_0^t \int_0^t h_2(\tau_1, \tau_2) x(t-\tau_1) x(t-\tau_2) d\tau_1 d\tau_2 \\
 & + \int_0^t \int_0^t \int_0^t h_3(\tau_1, \tau_2, \tau_3) x(t-\tau_1) x(t-\tau_2) x(t-\tau_3) \\
 & d\tau_1 d\tau_2 d\tau_3 + \dots
 \end{aligned} \tag{1}$$

식(1)을 이산형 다항식으로 표현하면 다음 식과 같은 Kolmogrov-Gabor의 다항식이 된다.

$$\begin{aligned}
 \Phi = & a_0 + \sum_i^N a_i x_i + \sum_i^N \sum_j^N a_{ij} x_i x_j + \sum_i^N \sum_j^N \sum_k^N a_{ijk} x_i x_j x_k \\
 & + \dots
 \end{aligned} \tag{2}$$

식(2)에서 입력은  $x_i = x(t-iT)$  이고 입력  $x(t)$  의  $i$  스텝 지연된 입력이며  $T$ 는 샘플링 시간(sampling time)이다. 식(2)의 회귀추정 문제에서 모델 구조를 모를 경우 입력계수  $N$ 이 4인 경우만 하더라도 추정해야 할 계수는 70 개항이 되고, 그에 따라 가능한 중회귀 구조의 수는 매우 많아진다[10]. 이와 같은 계산상의 어려움을 피하기 위해 모델구조를 자기 조직화(self organizing)하는 GMDH 알고리즘이 제시되었다.

시스템의 입력변수가  $x_1, x_2, x_3, \dots, x_N$  의  $N$  개일 때 예측 모델의 출력을  $y_m$  이라고 하면 예측 모델을 위한 일반 함수  $f$ 를 완전표현식이라 하고 식(3)과 같이 나타낸다.

$$y_m = f(x_1, x_2, x_3, \dots, x_N) \tag{3}$$

일반적인 회귀 모델과 GMDH 알고리즘 모델의 가장 큰 차이점은 완전표현식  $f$ 의 구성을 결정하는 데 있어서 입력 데이터 간의 연결 강도와 취사선택을 자기 조직화 원리에 의하여 수행한다는 것이다.

기본적인 GMDH 알고리즘은 식 (2)에서 2변수 2차 다항식을 선택하는 경우이며 식(4)와 같이 나타낸다.

$$z_k = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{4}$$

식 (4)와 같은 부분표현식으로 여러 층을 거듭하여 각 층에서 최적인 부분표현식을 선택함으로써 원하는 대상의 식별모델을 찾아내며, 이 과정을 통한 완전표현식 식 (3)을 자기 조직화 원리에 의하여 다음 순서로 결정한다.

단계 1. 임의의 2개의 입력변수의 조합으로 만들어진 아래의 식들을 시스템의 출력  $y_p$  에 근사하도록 결정한다.

$$\begin{aligned}
 z_{11} = G(x_1, x_2) \quad , \quad z_{12} = G(x_2, x_3) \quad , \dots \\
 z_{1S} = G(x_{N-1}, x_N)
 \end{aligned} \tag{5}$$

단계 2. 앞 단계의  $z_{11}, z_{12}, \dots, z_{1S}$  를 입력으로 하여 아래의 식들을 시스템의 출력  $y_p$  에 근사하도록 결정한다.

$$\begin{aligned}
 z_{i1} = G(z_{11}, z_{12}) \quad , \quad z_{i2} = G(z_{12}, z_{13}) \quad , \dots \\
 z_{iP} = G(z_{1,S-1}, z_{1S})
 \end{aligned} \tag{6}$$

여기서  $i$ 는 단계 2의 반복 횟수에 1을 더한 값이다.

단계 3. 단계 2에서 구한 함수  $G$ 를 부분표현식이라 부르고 이 식을 검사 데이터인 실제의 출력과 식 (7)에 의해 오차를 계산하여 오차가 적은 순으로  $M$ 개의 부분표현식을 선택한다. 통상  $M$ 은 입력변수  $x_N$ 의 개수  $N$ 보다 적게 한다.

$$r_j^2 = \frac{\sum_{i=nt+1}^n (y_p - z_n)^2}{\sum_{i=nt+1}^n y_i^2}, \quad j = 1, 2, \dots, \binom{m}{2} \tag{7}$$

여기서  $nt+1$ 은 검사 데이터의 시작 시점이고,  $n$ 까지의 합은 데이터 검사의 결과를 뜻하며  $z_{ij}$ 는  $i$ 번째 데이터들로  $j$ 번째 변수 조합으로 회귀분석한 결과치이다. 이 단계에서 얻은 부분표현식 중에서  $r_j < R$  인 것만 다

음 단계의 입력 변수로 활용된다. ( $R$ 은  $r_j$ 의 허용한계치이며 사전에 사용자가 정의 한다.)

단계 4. 단계 3에서 얻은  $r_j$ 의 최소값인  $RMIN_s$ 를 전 단계의  $RMIN_{s-1}$ 과 비교하여 에러가 커질 경우 혹은 더 이상의 개선이 일어나지 않을 경우 계산을 중지하고 단계 5로 진행하며, 그렇지 않을 경우 단계 2부터 다시 반복된다. 이 과정을 그림 3에 나타내었다.

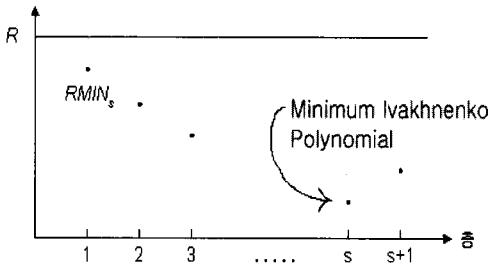


그림 3. GMDH 알고리즘의 종료 방법.  
Fig. 3. Stopping method of GMDH algorithm.

이 단계에서 계산된 식을 최종 식별된 완전표현식  $y_n$ 으로 정하며 최초 단계의 입력변수  $x_1, x_2, x_3, \dots, x_N$ 의 다차 다항식으로 Ivakhnenko 다항식[11]이라고 하며 식 (8)과 같다.

$$y_n = a + \sum_{i=1}^m b_i x_i + \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_i x_j + \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N a_{ijk} x_i x_j x_k + \dots \quad (8)$$

단계 5. 단계 4까지의 GMDH 알고리즘이 종료되면 자기 구성적 GMDH 모델이 완성된다. 그림 2와 같이 완성된 GMDH 모델의 출력과 실제의 출력을 비교하여 그 에러 값을 피드백하여 식 (9)와 같이 표현되는 각 노드의 파라미터들을 재조정한다.

$$w = [A \ B \ C \ D \ E \ F] \quad (9)$$

파라미터 조정을 위해서 다음과 같이 LMS 법칙을 적용한다.

$$w(k+1) = w(k) + \rho(y_p(k) - y_n(k))x(k), \quad k = 1, 2, \dots \quad (10)$$

여기서  $y_p(k)$ 는 실제의 출력이고  $y_n(k)$ 는 예측된 출력이며

$$x(k) = [1 \ x_i \ x_j \ x_i x_j \ x_i^2 \ x_j^2] \quad (11)$$

이다.

### III. 시뮬레이션 결과

피드백에 의한 GMDH 알고리즘과 전형적인 알고리즘을 전력 수요 예측에 적용하여 그에 따른 에러율 차이를 검토하고자 한다. 전력 수요량을 예측하기 위하여 1984년부터 1998년까지 전력 수요량에 영향이 가장 많은 GDP, 수출액, 수입액, 경제활동 인구, 취업자 수, 석유 소비량, 평균 기온을 입력으로 사용하였다.

장기 예측의 경우에 대하여 제안한 GMDH 알고리즘의 예측 정확도를 입증하기 위해서 1984년부터 1993년까지 10개년을 입력 데이터로 하여 1999년부터 2008년까지 10개년의 전력 수요를 예측하였다. 그림 4와 5에서 각각 전형적인 GMDH 알고리즘과 피드백에 의한 GMDH 알고리즘을 시뮬레이션한 결과를 그래프로 나타내었다.

전형적인 GMDH 알고리즘의 경우 평균 에러율이 4.29%였으며 피드백에 의한 GMDH 알고리즘의 경우는 2.77%였다. 따라서 본 논문에서 제안한 피드백에 의한 GMDH 알고리즘에 의해 더 정확한 예측 결과를 얻을 수 있다는 것을 알 수 있었으며, 본 시뮬레이션의 경우에는 35.4%의 예측 성능 개선의 결과를 얻을 수 있었다.

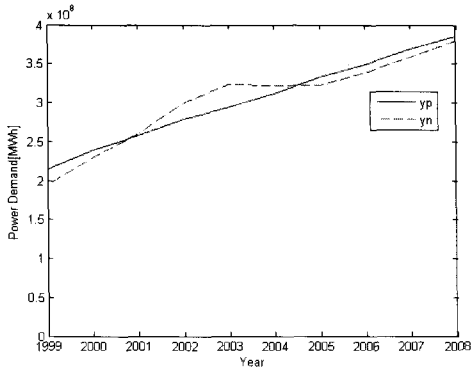


그림 4. 전형적인 GMDH 알고리즘의 전력 수요 장기 예측.

Fig. 4. Power demand long term forecasting of conventional GMDH algorithm.

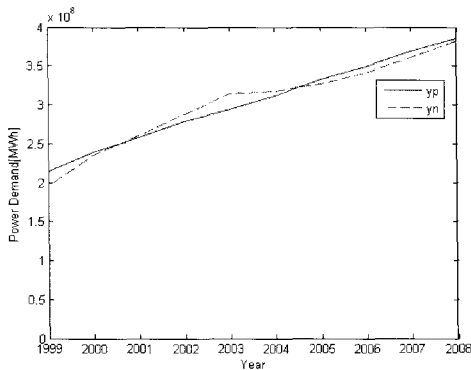


그림 5. 피드백에 의한 GMDH 알고리즘의 전력 수요 장기 예측.

Fig. 5. Power demand long term forecasting of GMDH algorithm by feedback.

단기 예측의 경우에도 제안한 GMDH 알고리즘의 예측 정확도를 입증하기 위해서 1989년부터 1998년까지 10개년을 입력 데이터 기간으로 하여 1999년부터 2008년까지 10개년을 예측하였다. 그림 6과 7에서 각각 전형적인 GMDH 알고리즘과 피드백에 의한 GMDH 알고리즘을 시뮬레이션한 결과를 그래프로 나타내었다.

전형적인 GMDH 알고리즘의 경우 평균 에러율이 1.62%였으며 피드백에 의한 GMDH 알고리즘의 경우는 0.99%였다. 따라서 단기 예측의 경우에도 본 논문에서 제안한 피드백에 의한 GMDH 알고리즘에 의해 더 정확

한 예측 결과를 얻을 수 있다는 것을 알 수 있었으며, 본 시뮬레이션의 경우에는 38.9%의 예측 성능 개선의 결과를 얻을 수 있었다.

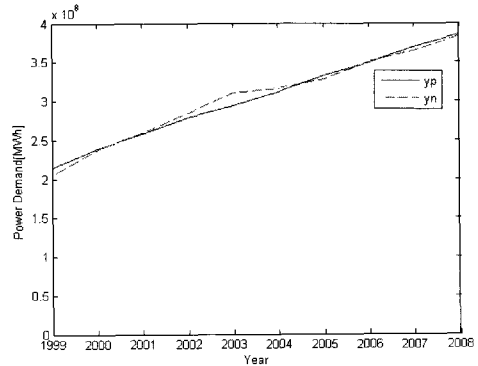


그림 6. 전형적인 GMDH 알고리즘의 전력 수요 단기 예측.

Fig. 6. Power demand short term forecasting of conventional GMDH algorithm.

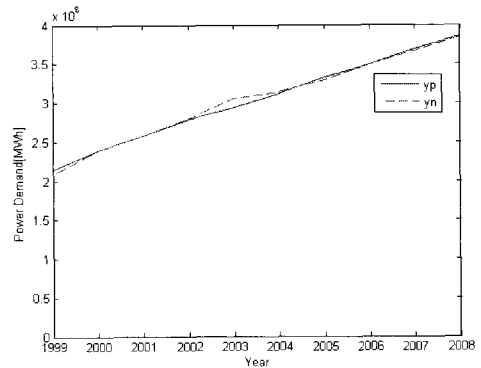


그림 7. 피드백에 의한 GMDH 알고리즘의 전력 수요 단기 예측.

Fig. 7. Power demand short term forecasting of GMDH algorithm by feedback.

#### IV. 결론

본 논문에서는 GMDH 알고리즘에 피드백 기능을 추가하고 LMS 학습 법칙을 적용하여 노드의 연결 강도를 조정하는 알고리즘을 제안하였다. 전형적인 GMDH 알고리즘이 층별로 최적의 구조가 결정되지만 전체적으

로는 최적의 구조가 결정되지 않을 수도 있다는 문제점을 해결하기 위해 GMDH로 예측된 출력을 실제의 출력과 비교하여 그 에러를 피드백하여 전체적으로 최적의 구조를 가지는 GMDH 예측 모델을 구성함으로써 보다 정확한 예측이 가능하도록 하였다.

본 논문의 시뮬레이션 결과에 따르면, 전력 수요 예측의 결과에서 전형적인 GMDH 알고리즘을 이용한 예측 결과의 평균 에러율은 장기 예측과 단기 예측의 경우에 각각 4.29%와 1.62%이고, 피드백에 의한 GMDH 알고리즘을 이용한 예측 결과의 평균 에러율은 장기 예측과 단기 예측의 경우에 각각 2.77%와 0.99%였다. 즉, 피드백을 적용한 알고리즘의 결과가 장기 예측과 단기 예측에 대해서 각각 35.4%와 38.9% 개선되었음을 알 수 있었다. 따라서 본 논문에서 제안한 피드백에 의한 GMDH 알고리즘을 적용하여 여러 분야의 GMDH 예측에 적용한다면 보다 정확한 예측이 가능할 것으로 기대된다.

### 참고문헌

[ 1 ] H. Tamura and T. Kondo, "Heuristics free group method of data handling algorithm of generating optimal partial polynomials with application to air pollution prediction," *Int. J. Syst. Sci.*, vol. 11, pp. 1095-1111, 1980.

[ 2 ] A. G. Ivakhnenko and G. A. Ivakhnenko, "The review of problems solvable by algorithms of the GMDH, *Pattern Recognition and Image Analysis*, vol. 5, no. 4, pp. 527-535, 1995.

[ 3 ] M. S. Voss and X. Feng, *Emergent system identification using particle swarm optimization*. Hutchinson, 2002.

[ 4 ] X. H. Tong, J. C. Kuang, X. Y. Wang, and T. X. Qi, "Setting up prediction model of gas well prediction rate by various methods," *Natural Gas Industry*, vol. 16, no. 6, pp. 49-53, 1996.

[ 5 ] F. J. Chang and Y. Y. Hwang, "A self-organization algorithm for real-time flood forecast," *Hydrological Processes*, vol. 13, pp. 123-138, 1999.

[ 6 ] W. Liu, X. Huang, G. Wang, and T. Chai, "The modeling and application of cost prediction based on neural network," *5th Asian Control Conference*, pp. 1308-1312, 2004.

[ 7 ] I. G. Damousis, M. C. Alexiadis, J. B. Theocharis, and P. S. Dokopoulos, "A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation," *IEEE Trans. Energy Conversion*, vol. 19, no. 2, pp. 352-361, June 2004.

[ 8 ] N. Y. Nikolaev and H. Iba, "Polynomial harmonic GMDH learning networks for time series modeling," *Neural Networks*, vol. 16, pp. 1527-1540, 2003.

[ 9 ] S.-P. Lo and Y.-Y. Lin, "Application of abductive polynomial network and grey theory to drill flank wear prediction," *Applied Artificial Intelligence*, vol. 18, pp. 713-733, 2004.

[10] 안기범, "GMDH를 이용한 모델링에 관한 연구," 동아대학교 대학원 석사학위 논문, 1995년.

[11] A. G. Ivakhnenko and A. A. Zhohnarskiy, "Estimating the coefficients of polynomials in parametric GMDH algorithms by the improved instrumental variables method," *Soviet Automatic Control*, vol. 25, no.3, pp. 25-32, 1992.

### 저자소개

홍연찬(Yeon Chan Hong)

한국해양정보통신학회 논문지  
제 13권 제 6호 참조