

---

# GMDH를 이용한 비선형 시스템의 모델링 성능 개선

홍연찬\*

Performance Improvement of Nonlinear System Modeling Using GMDH

Yeon Chan Hong\*

---

본 논문은 2009년도 인천대학교 학술연구조성비 지원에 의하여 수행되었음

---

## 요 약

비선형 동적 시스템을 모델링하기 위해 GMDH(Group Method of Data Handling)를 적용한 많은 연구들이 수행되어 왔다. 그러나 모델링의 정확성을 위해서는 계산량이 크게 증가한다. 그러므로 본 논문에서는 입력 데이터를 취사선택하는 기준을 점감적으로 조정함으로써 적어도 정확성을 유지하면서 전형적인 GMDH의 단점인 과도한 계산을 피할 수 있는 방법을 제안한다. 컴퓨터 시뮬레이션 결과, GMDH 알고리듬의 계산량을 성공적으로 줄일 수 있었고 에러율도 소폭 줄일 수 있었다.

## ABSTRACT

There have been many researches applying GMDH for modelling nonlinear dynamic systems. However, these methods require a great amount of computation in return of the accuracy. Thus, in this paper, we propose a method to reduce the amount of computation in GMDH by adjusting the adopting criterion of input data in decrement while at least maintaining the accuracy. The simulation result verifies that the proposed method can successfully reduce the amount of computation without the expense of the error rate, if not significantly better.

## 키워드

비선형 시스템, GMDH 알고리듬, 모델링, 계산량, 취사선택 기준, 에러율

## Key word

nonlinear system, GMDH algorithm, modelling, amount of computation, adopting criterion, error rate

---

\* 인천대학교(ychong@incheon.ac.kr)

접수일자 : 2010. 04. 06

심사완료일자 : 2010. 06. 29

## I. 서 론

시스템의 수학적 모델링은 특성이 비선형이고 차수가 크기 때문에 일반적으로 난해하고 복잡하다. 이러한 복잡한 시스템들에 대하여 시스템의 인식 문제를 해결하기 위한 많은 연구들이 이루어졌다. 신경회로망, 퍼지 이론, SVM(support vector machine), GMDH 등과 같은 인공지능 기법을 적용한 예측이 국내외적으로 활발히 연구되고 있다[1,2]. 특히, 예측모델의 구성이 용이하고 기존의 시계열 해석법이나 회귀분석법에 비해 우수하다고 평가되고 있는 신경회로망 기법이 주류를 이르고 있으며, 여러 가지 모델이 제시되고 있다. 그러나 신경회로망을 이용한 수요예측 기법은 모든 데이터를 처리해야 하는 단점이 있기 때문에 여러 가지 요인을 데이터로 사용하는 전력수요 예측에서는 장기적인 예측에 단점을 보이고 있다. 그러므로 기존 방법의 단점을 보완 할 수 있는 방법에 대한 연구가 요구되고 있다[3]. GMDH는 각종에서 유용한 데이터를 자체적으로 추출할 수 있는 구조의 다층 알고리듬이므로 구조적 효율성이 높다고 할 수 있다[4-7].

전력설비 건설 계획 수립에는 국가 차원의 에너지 안보, 차원배분, 환경정책과 전력의 수요관리, 예비전력 수준 등 많은 고려 요소가 존재한다. 그중에서도 안정적이고 양질의 전력을 공급하는 것이 공익산업으로서 전력 산업이 설정하는 기본 목표라 할 때, 전력설비 투자를 입안, 계획하는 단계에서부터 전력 수요에 대한 예측은 필수 불가결한 요소이다.

과거의 전력 수요에 사용되어온 방법은 최대 전력 수요를 예측하는 전단계로서, 수요전력량에서 일부하곡선을 작성하여 최대 전력수요가 발생하는 일부하를 예측하는 부하곡선 방식과 연간 수요전력량과 경제성장률에 의한 산업 구조, 공장의 조업도, 휴일일수, 계절 수요의 가중도, 부하율이 다른 전력의 구성비 등을 고려한 연부하율 관계에서 장기 최대 전력 수요를 예측하는 연부하율 방식을 많이 사용했고, 최근에는 소득효과와 가격 효과를 포함하는 일반적인 수요함수의 형태에 기상요인을 고려한 시계열 모형을 주로 사용한다[8].

그러나 시계열 모형은 모형이 선형이라는 가정 하에 전력 예측을 하기 때문에 전력 수요 예측에 영향을 미치

는 요소가 다양하고 복잡한 비선형인 경우 모형의 결과가 신뢰를 잃게 된다.

이런 단점을 극복하고자 단기 전력 수요 예측에 GMDH 알고리듬을 적용한 결과 효율적인 활용성과 정확성을 보였다[9].

그러나 급변하는 세계정세와 장기 불황 속에서 장기적인 안목으로 전력 수요 예측의 필요성이 언제 보다도 높게 제기되고 있고, 많은 예측 요소로 인한 데이터의 증가와 높은 비선형성, 이로 인한 계산량의 증가로 효율성의 저하는 높은 성능하의 컴퓨터 환경 속에서도 아주 중요한 문제로 대두되고 있다.

본 연구에서는 좀 더 적은 계산량과 높은 예측 정확성을 얻기 위하여 시계열 예측에서 우수한 성능을 보인 GMDH 알고리듬을 이용하였다. GMDH 알고리듬에서 데이터를 취사선택하는 기준을 변화시킴으로써 높은 예측률과 계산량을 감소시키는 방법에 대해 제안한다. 또한 이를 장기 전력 수요 예측에 적용하여 에러율의 변화와 계산량의 변화를 확인하여 알고리듬을 검증한다.

## II. 수정된 GMDH 알고리듬

일반적으로 차분 방정식(difference equation)에 의한 비선형 다이내믹 시스템의 상태 방정식은 다음과 같다.

$$\begin{aligned} x(k+1) &= \Phi[x(k), u(k)] \\ y(k) &= \Psi[x(k)] \end{aligned} \quad (1)$$

이때,  $x(k)$ ,  $y(k)$ ,  $u(k)$ 는 각각 상태, 출력, 입력을,  $\Phi[\cdot]$ 와  $\Psi[\cdot]$ 는 비선형 함수를 나타낸다. 이와 같은 비선형 차분 방정식을 재구성하면 다음과 같은 식으로 나타낼 수 있다.

$$\begin{aligned} y(k+1) &= f[y(k), y(k-1), \dots, y(k-n+1), \\ &\quad u(k), \dots, u(k-m+1)] \end{aligned} \quad (2)$$

이때  $f[\cdot]$ 는 비선형 함수이다.

GMDH 알고리듬은 입력 노드(input node)들의 집합으로 이루어져 있으며, 이 입력 노드들로부터 출력을 얻기 위하여 중간층들을 거치게 된다. 각 층에서의 노드들은 바로 전 단계 층의 노드들의 출력들로부터 두개씩을 취하여 출력을 만들어낸다[10].

GMDH 알고리듬의 구조를 그림 1에 나타내었다. 그림 1에서  $z_{s,t}$ 는  $s$ 번째 층의  $t$ 번째 노드의 출력을 나타내며  $m$ 은 입력의 개수다. 두 번째 층은  $i$ 개의 노드들로 구성되어 있는데,  $i < \frac{m(m-1)}{2}$  이고  $j < \frac{i(i-1)}{2}$ 이다.

이 부등식들이 성립하는 이유는 뒤에서 언급할 데이터 선택 기준에 의해 영향이 적은 데이터를 버리기 때문이다. 최종적으로 하나의 출력을 얻게 되는 과정도 데이터 선택 기준의 최소값에 의해 결정된다. 모든 층의 각각의 노드들의 출력은 다음 식과 같이 2차 방정식의 형태를 취하게 된다.

$$z_{s,t} = a_{s,t} + z_{(s-1),u}^2 + b_{s,t}z_{(s-1),v}^2 + c_{s,t}z_{(s-1),u} + d_{s,t}z_{(s-1),v} + e_{s,t}z_{(s-1),u}z_{(s-1),v} + f_{s,t} \quad (3)$$

$$+ d_{s,t}z_{(s-1),v} + e_{s,t}z_{(s-1),u}z_{(s-1),v} + f_{s,t}$$

이때  $a_{s,t}, b_{s,t}, c_{s,t}, d_{s,t}, e_{s,t}, f_{s,t}$ 는 노드의 연결강도(connection weight)이고, 아래첨자 u와 v는 임의의 u번째 와 v번째 노드를 가리킨다.

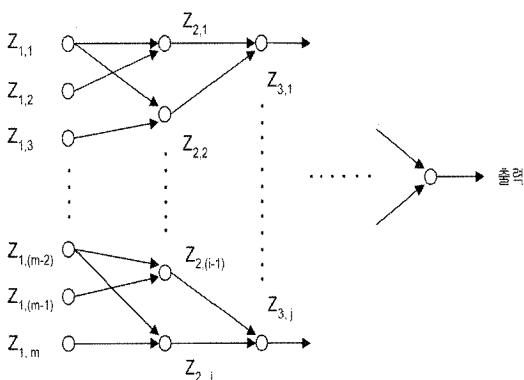


그림 1. GMDH 알고리듬의 구조.

Fig. 1. Structure of GMDH algorithm.

예를 들어 3번째 층으로 들어오는  $i$ 개의 입력들의 조합의 수( $C_2^i$ )에 의해서 다음 층의 노드 수가 결정되고 난 후, 이 노드들의 출력이 식(3)에 의해 각각 계산되며, 계산된 출력의 적합성 여부를 성능 기준에 의하여 판별한다는 것이다. 이처럼 각각의 층에서의 적합지 못한 노드들을 판별해내기 위해서는 식(4)를 이용하여 원하는 최종 출력과 실제 노드 출력들 사이의 에러를 계산하는 것이다.

$$r_{s,t} = |y(k) - (z_{s,t})_n| \quad (4)$$

여기서  $(z_{s,t})_n$ 은  $s$ 번째 층과  $t$ 번째 노드의 출력인  $z_{s,t}$  벡터의 마지막 요소인  $n$ 번째 요소를 의미한다. 이 에러의 크기를 구하여 큰 에러를 가진 노드는 없애고 에러가 작은 노드들만을 가지고 앞서 설명한 과정들을 원하는 출력을 얻을 때까지 반복하게 되는 것이다. 그리고 이 알고리듬은 일반적으로 사용되고 있는 신경망 구조와는 달리 중간층의 수와 노드들의 수가 미리 정해지는 것이 아니다.

본 논문에서 제안한 알고리듬을 GMDH 알고리듬[11]에 적용하여 다음과 같이 5 단계로 나누어 비선형 시스템을 모델링한다.

#### 단계 1. 변수벡터의 구성

식(3)에 의하여  $k$  순간에서  $s$ 층의  $t$ 번째 노드의 출력  $z_{s,t}(k)$ 를 만들어 내며 다음 식과 같이 간단하게 표현할 수 있다.

$$z_{s,t}(k) = X_{s-1}\theta_s \quad (5)$$

$$X_{s-1} = [z_{(s-1),u}^2(k) \ z_{(s-1),v}^2(k) \ z_{(s-1),u}(k) \ z_{(s-1),v}(k) \ z_{(s-1),u}(k) \ z_{(s-1),v}(k) \ 1]$$

$$\theta_s = [a_{s,t} \ b_{s,t} \ c_{s,t} \ d_{s,t} \ e_{s,t} \ f_{s,t}]^T \quad (6)$$

여기서  $X_{s-1}$ 은 차수가  $n \times 6$ 인 데이터 벡터로서 두 벡터 입력  $z_{(s-1),u}(k)$ 와  $z_{(s-1),v}(k)$ 를 합성한 것이다.

식(5)는  $\theta_s$ 와 함께 2차 방정식을 만들어낸다. 이 때  $z_{s,t}(k)$ 를 얻기 위해서는, 먼저 식(5)를 풀어 계수 벡터

인 연결강도  $\theta_s$ 를 얻어야 한다. 각 노드의 출력  $z_{s,t}(k)$  가 최종적으로 시스템의 출력  $y(k)$ 가 되도록 하는 것이 이 알고리듬의 목적이므로 그 일반식은 다음과 같다.

$$\theta_s = (X_{s-1}^T X_{s-1})^{-1} X_{s-1}^T y(k) \quad (7)$$

### 단계 2. 영향이 적은 데이터의 제거

앞서 설명했듯이, 단계 1에서 구해진 모든 노드들의 출력들이 다음 층으로 전달되는 것은 아니다. 즉 적자생존의 원칙을 적용해야 한다. GMDH 알고리듬에서는 출력과 새로 생성된 데이터를 훈련 부분과 검사 부분으로 나누어야 한다. 따라서 출력과 생성된 데이터의 훈련부분 과정에서 구해진 평균제곱에러에 의해 필요 없는 노드들을 제거한다.

$r_{s,t}$  값의 증가 순으로  $s$  층의 모든  $z_{s,t}(k)$ 를 다시 정렬시킨다. 그리고 기준 값  $R$ 보다 큰 값을 갖는 노드들을 제거한다. 그리고 제거되지 않은 변수  $z_{s,t}(k)$ 들을  $(s+1)$  층의 입력으로 사용한다.

### 단계 3. 최적성의 시험

이 단계에서는 훈련을 계속할지를 판별하기 위하여, 단계 2에서 계산된  $r_{s,t}$ 들의 최소값  $R\text{MIN}_s$ 를 전 단계의  $R\text{MIN}_{s-1}$ 과 비교하는 것이다.  $R\text{MIN}_s < R\text{MIN}_{s-1}$  이면 단계 1과 단계 2를 반복하고,  $R\text{MIN}_s \geq R\text{MIN}_{s-1}$  이면  $R\text{MIN}_{s-1}$ 에 해당하는  $z_{(s-1),t}(k)$ 를 출력의 추정치  $\hat{y}(k)$ 로 보고 모든 훈련을 끝내게 되는 것이다. 그러면 이  $z_{(s-1),t}(k)$ 는 Ivakhnenko polynomial이 된다.

### 단계 4. $R\text{MIN}_s$ 의 순차적 감소 적용

단계 3에서와 같이 일반적인 GMDH 알고리듬에서는 예상이 기준  $R$ 보다 작은 노드들이 모두 선택되어 층이 증가됨에 따라서  $R\text{MIN}_s$ 와  $R$  사이의 간격이 커지므로 노드의 수가 크게 증가하여 계산량의 과다로 이어진다.

이런 단점을 극복하고자  $R\text{MIN}_s$ 와  $R$  사이의 간격을 순차적으로 감소시킴으로써 계산량의 과다를 방지할 수 있도록 하는 수정된 GMDH 알고리듬의 기준 값  $R_s$ 는 다음과 같다.

$$R_s = R\text{MIN}_s + \delta \quad (8)$$

여기서  $\delta$ 는 미리 정한 일정한 값이다.

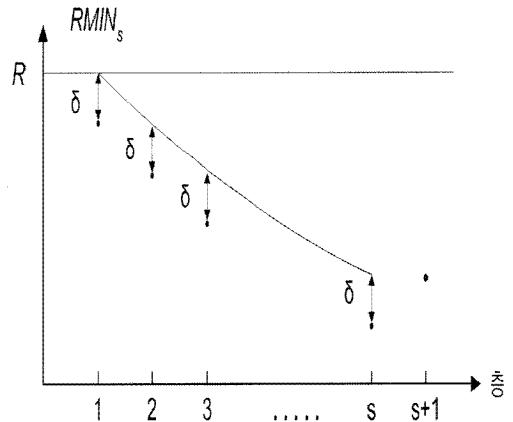


그림 2. 수정된 GMDH 알고리듬의 종료법.  
Fig. 2. Finish method of modified GMDH algorithm.

### III. 점감적 감소 취사선택 기준

단계 5. 위의  $R_s$ 의 순차적 감소를 이용한 GMDH 알고리듬은 단기 예측에 적용한 결과 높은 성능을 보여주었다. 그러나 장기 예측을 목적으로 하는 경우에는 많은 입력 노드의 증가로 인한 계산량의 증가가 대두될 수 있다.

위와 같은 문제점을 개선하기 위하여 가중치를 두어  $R_s$ 를 감소시키는 점감적 GMDH 알고리듬을 다음과 같이 제안한다.

$$R'_s = R\text{MIN}_s + \delta' \quad (9)$$

$$\delta' = \delta - (n-1)\beta \quad (10)$$

식 (9)와 (10)에 의해 일정한 값  $\delta$ 에 세대마다 가중치  $\beta$ 를 주어 그림 3과 같이  $R\text{MIN}_s$ 와  $R$  사이의 간격을 세대가 증가함에 따라 더 감소시켜서 새로운 기준 값  $R'_s$

를 만들어 내는 새로운 방식을 제안한다.

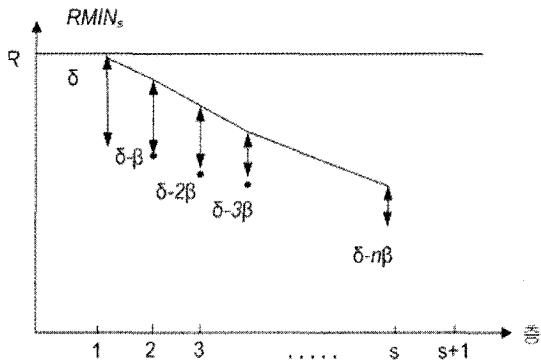


그림 3. 점감적 GMDH 알고리듬의 종료법.  
Fig. 3. Finish method of decremental GMDH algorithm.

#### IV. 시뮬레이션 결과

본 논문은 1991년 1분기부터 1997년 4분기까지의 전력수요량, GDP, 수출, 수입, 취업자, 경제활동 인구, 석유소비량, 평균 기온의 training data를 두고 1998년 1분기부터 2003년 4분기까지의 예측 결과를 구하기 위하여, 장기 전력 수요 예측에 test data를 두어 본 논문에서 제안한 GMDH 알고리듬을 검증하였다.

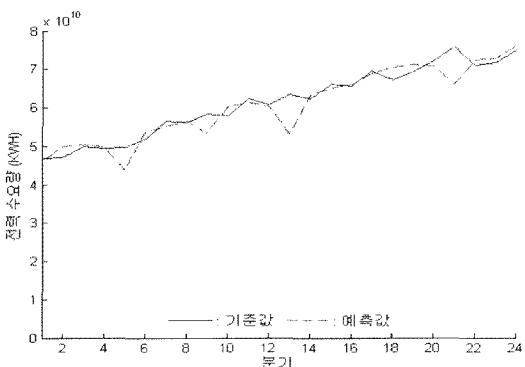


그림 4.  $\beta=0$ 일 때의 전력 수요량 예측 결과.  
Fig. 4. Electric power demand forecasting result when  $\beta=0$ .

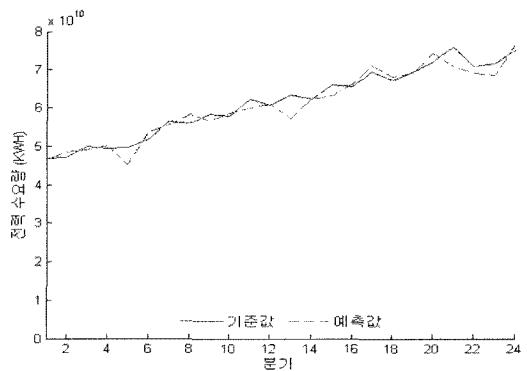


그림 5.  $\beta=0.06$ 일 때의 전력 수요량 예측 결과.  
Fig. 5. Electric power demand forecasting result when  $\beta=0.06$ .

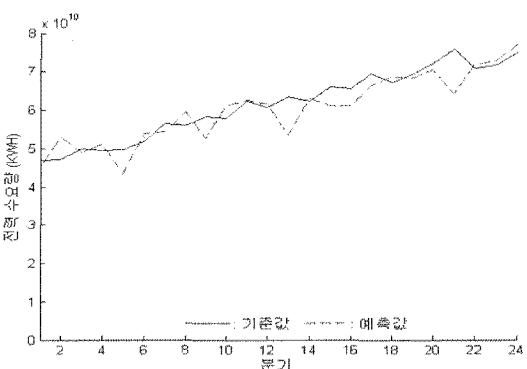


그림 6.  $\beta=0.1$ 일 때의 전력 수요량 예측 결과.  
Fig. 6. Electric power demand forecasting result when  $\beta=0.1$ .

제안된 GMDH 알고리듬을 시뮬레이션하여  $\delta$ 와  $\beta$ 에 따른 에러율과 계산량 변화를 확인하였다. 그림 4, 그림 5, 그림 6, 그림 7은  $\delta=2$ 로 놓고  $\beta$ 를 변화시키면서 전력 수요량 예측을 시뮬레이션한 결과이다.

표 1을 보면  $\beta=0$ 일 때 계산량이 크고 평균 에러율은 2.98 %로 높게 나왔다. 여기서 계산량은 전체 노드의 개수를 나타낸다.  $\beta=0.01$ 부터  $\beta=0.03$ 까지는 평균 에러율은 지속적인 감소를 보였고  $\beta=0.04$ 부터  $\beta=0.06$ 일 경우 가장 적은 에러율을 나타내었으며 계산량도 지속적으로 감소하였다.  $\beta=0.07$ 부터  $\beta=0.09$ 는 계산량은 감소되고 있지만 다시 에러율이 증가하였고  $\beta=0.1$ 에서는 계산

량이 감소하였지만 에러율이 최대가 됨을 알 수 있다. 그 이유는  $\beta$  값이 커지면서 기준 값이 작아져서 노드를 지나치게 제거시켰기 때문이다.

위의 결과를 바탕으로 에러율이 낮으면서도 계산량이 적어서 가장 효율적인 경우가  $\beta=0.06$ 임을 알 수 있다. 계산량의 감소는 (9)와 (10)식의 기준 값을 조정하여 유용한 데이터들만을 추출함으로써 이루어졌고, 유용한 데이터들에 의한 training에 의해서 예측의 정확도가 유지될 수 있었다.

표 1.  $\beta$ 에 대한 전력 수요량 예측 결과  
Table 1. Electric power demand forecasting result  
for  $\beta$

$\beta$	계산량	평균 에러율
0	23476	2.98
0.001	23434	2.96
0.01	23434	2.96
0.02	23434	2.96
0.03	14423	2.91
0.04	10836	2.81
0.05	5947	2.81
0.06	2932	2.81
0.07	1699	3.33
0.08	455	3.60
0.09	325	3.80
0.1	214	3.94

## V. 결론

본 논문에서는 GMDH 알고리듬을 이용하여 장기 전력 수요 예측을 수행하였다. 전력설비 건설은 장기적인 기간이 소요되는 사업이다. 따라서 정확한 장기 전력 수요 예측이 필요하다. 또한 장기 전력 수요 예측에 영향이 미치는 요소는 GDP, 수출, 수입, 취업자, 경제 활동 인구, 석유소비량과 같은 경제적인 요소와 기후적인 요소인 기온을 선정하여 적용하였다. 이런 다양하고 복잡한 요소들은 비선형적인 특성을 보인다. 즉 비선형 시스템의 수학적 모델링의 가장 큰 단점은 계산량의 증가이다.

이런 단점을 극복하고자 장기 전력 수요 예측에 본 논문에서 제안한 향상된 GMDH 알고리듬을 적용하였다.

그 결과  $\delta$ 와  $\beta$ 를 적용한 경우 계산량과 평균 에러율을 감소시킬 수 있었다.  $\delta$ 와  $\beta$ 를 동시에 적용한 GMDH 알고리듬의 경우  $\delta$ 만 적용한 경우보다 계산량은 87.5% 줄일 수 있었고 평균 에러율은 5.7% 줄일 수 있었다. 따라서 본 논문에서 제안한 GMDH 알고리듬을 장기 전력 수요 예측에 적용한 결과 정확하고도 빠른 예측이 가능하였다.

## 참고문헌

- [1] S. S Pappas and L. Ekonomou, "Comparison of artificial intelligence methods for predicting the time series problem," Proc. of the 6th WSEAS Int. Conf. on Simulation, Modeling and Optimization, Lisbon, Portugal, pp. 35-40, Sept. 2006.
- [2] T. Kondo, "A new algorithm of multi-layered neural network using heuristic self-organization," Proc. of the 17th Int. Symp. on Mathematical Theory of Networks and Systems, Kyoto, Japan, pp. 911-916, July 2006.
- [3] 이창호, 채승용, "도서지역 전력수요예측모형 연구," 한국 에너지 공학 학회 추계 학술 논문집, pp. 271-280, 2002.
- [4] T. Yoshimura, R. Kiyozumi, K. Nishino, and T. Soeda, "Application of revised GMDH algorithms to the prediction of air pollutant concentrations," Proc. 23rd IFAC World Congr., vol. XXIII, pp. 13-17, Kyoto, Japan, Aug. 1981.
- [5] Y. Nishikawa, J. Nomura K. Sawada, and T. Kurio, "A Method for Predicting Sales Amount by Use of IWSM and GMDH," System and Control, Vol.27, No.7, pp.475-479, 1983.
- [6] 안기범, "GMDH를 이용한 모델링에 관한 연구," 동아대학교 대학원 석사 학위 논문, 1995.
- [7] N. Ponomareva, M. Alexandrov, A. Gelbukh, "Performance of inductive method of model self-organization with incomplete model and noisy data," 2008 7th Mexican Int. Conf. on Artificial

- Intelligence, pp. 101-108, 2008.
- [ 8 ] 이창복, "GMDH 알고리즘을 응용한 전처리 선별과 인공신경망을 이용한 한국종합주가지수 예측에 관한 연구," 연세대학교 대학원 석사 학위 논문, 1999.
- [ 9 ] 박성희, "퍼지 신경회로망을 이용한 장기 전력수요 예측," 서울대학교 대학원 석사 학위 논문, 1996.
- [10] S. J. Farlow, Self-Organizing Methods in Modeling / GMDH Algorithms: Marcel Dekker Inc., 1984.
- [11] 이동철, 홍연찬, "GMDH를 이용한 전력 수요 예측 알고리듬 개발," 퍼지 및 지능시스템학회 논문지, 제 13권, 제 3호, pp. 360-365, 2003년 6월.

### 저자소개

홍연찬(Yeon Chan Hong)

한국해양정보통신학회 논문지  
제 13권 제 6호 참조