

GMDH를 이용한 전력 수요 예측 알고리즘 개발

Development of Power Demand Forecasting Algorithm Using GMDH

이동철*, 홍연찬**

Dong Chul Lee and Yeon Chan Hong

* 액트통신기술 (주), ** 인천대학교 전자공학과

요 약

본 논문에서는 데이터의 효율적인 활용과 정확성에서 보다 우수한 특성을 보이는 GMDH(Group Method of Data Handling) 알고리즘을 전력수요예측에 적용함으로써 입력 데이터의 선정을 용이하게 하였고, 다양한 데이터를 기반으로 보다 정확한 예측을 할 수 있게 하였다. 그리고, 예측 시에 경제적인 요인(GDP, 수출, 수입, 취업자 수, 경제활동인구, 석유소비량)과 기후적인 요인(평균기온)을 모두 고려하였다. 또한 목표 예측 기간을 1999년 1/4분기에서 2001년 1/4분기까지 9개의 분기로 가정하고, 가정한 목표 기간의 예측 정확도를 높이기 위해 3단계의 시뮬레이션 과정(최적 입력 분기 수를 결정하는 과정, 입력 데이터와 예측값의 시간적 연관성을 분석하는 과정, 입력 데이터의 최적화 과정)을 이용함으로써 더 정확한 전력수요예측 방법을 제시하였고, 제안된 기법으로 목표한 예측 기간에서 0.96%의 평균 에러율을 얻을 수 있었다.

Abstract

In this paper, GMDH(Group Method of Data Handling) algorithm which is proved to be more excellent in efficiency and accuracy of practical use of data is applied to electric power demand forecasting. As a result, it became much easier to make a choice of input data and make an exact prediction based on a lot of data. Also, we considered both economy factors(GDP, export, import, number of employee, number of economically active population and consumption of oil) and climate factors(average temperature) when forecasting. We assumed target forecast period from first quarter 1999 to first quarter 2001, and suggested more accurate forecasting method of electric power demand by using 3-step computer simulation processes(first process for selecting optimum input period, second for analyzing time relation of input data and forecast value, and third for optimizing input data) for improvement of forecast precision. The proposed method can get 0.96 percent of mean error rate at target forecast period.

Key Words : GMDH, electric power demand forecasting, computer simulation, optimizing input data

1. 서 론

본 논문에서 이용한 GMDH(Group Method of Data Handling) 알고리즘에 관한 현재의 연구동향은 다음과 같다. Yoshimura는 부분표현식의 선택에 AIC(Akaike's information criterion)를 이용한 개량형 GMDH 알고리즘을 이용하여 산소, 농도, 풍속, 풍향을 입력으로 하여 공기 오염 농도의 예측 모델을 구성하였다[1]. Nishikawa는 기업에 있어서 이전의 판매 실적과 평균값을 이용하여 판매 예측 모델을 구하여 생산, 판매, 재고 계획업무에 GMDH 알고리즘을 적용하였다[2]. Ann은 화력발전량과, 원유가, 환율을 독립변수로 이용하여, 석유소비량을 예측하는데 GMDH 알고리즘을 이용하였다[3]. Lee는 주가예측의 정확성을 높이기 위하여 GMDH 알고리즘을 이용함으로써 주식투자 전략을 위한

의사 결정과정에 보다 더 유용한 정보를 제공할 수 있도록 하였다[4].

이와 같이 GMDH 알고리즘은 어떠한 비선형 시스템에 대한 모델링과 예측에 강점을 보이고 있다. 따라서, 본 연구에서는 여러 가지 요인에 의하여 비선형적인 변동 특성을 보이는 전력 수요량의 예측에 이용하였다.

전력은 오늘날 급격하게 성장하고 산업활동이 활발한 현대 산업 사회에서 산업의 원동력이 되고, 국민생활의 가장 필수적인 요소이다. 따라서 전력계통에 최적 계획 및 경제적인 운용 기법을 적용함으로써 산업활동과 문화생활에 있어서 얻어지는 이득이 막대하기 때문에 전력수요를 사전에 정확하게 예측하는 것은 중요한 일이다.

전력수요는 기상요인이나 사회·경제적인 요인에 의해 불규칙하게 변동되고, 주기적으로 변동하는 특성이 있어 정확하게 예측하기 어렵다. 현재까지 대부분의 전력수요예측에서 사용되어온 방법은 시계열 해석방법이나 회귀분석법을 이용한 형태가 주류를 이루고 있으며 최근 시계열 해석법이 갖는 기본적인 한계를 극복하고자 신경회로망, 퍼지이론 등과 같은 인공지능 기법을 적용한 예측이 국내외적으로 활발히 연구되고 있다. 특히, 예측모델의 구성이 용이하고 기존의 시계

접수일자 : 2002년 5월 10일

완료일자 : 2003년 5월 23일

본 연구는 2001년도 인천대학교 학술연구조성비 지원에 의하여 수행되었음

열해석법이나 회귀분석법에 비해 우수하다고 평가되고 있는 신경회로망 기법이 주류를 이루고 있으며 여러 가지 모델이 제시되고 있다. 그러나, 신경회로망을 이용한 수요예측기법은 모든 데이터를 처리해야하는 단점이 있기 때문에 여러 가지 요인을 데이터로 사용하는 전력수요예측에서는 장기적인 예측에는 단점을 보이고 있다. 그러므로, 기존 방법의 단점을 보완할 수 있는 방법에 대한 연구가 요구되고 있다[5~8].

본 논문에서는 기존의 예측 기법에 비하여 Data의 효율적인 활용과 정확성에서 보다 우수한 특성을 보이는 GMDH 알고리즘을 전력수요예측에 적용하고자 한다[9~13]. GMDH 알고리즘은 신경망과 달리 매 셀마다 적자생존원칙을 적용하여 필요한 셀들만 모형을 진행하기 때문에 상대적으로 더 많은 데이터를 처리 할수 있다는 장점이 있다[14~18]. 그러므로 경제, 사회 기후적 요인을 모두 적용하여 전력수요예측을 가능하게 할 것으로 예상된다.

본 논문에서는 많은 데이터를 처리하기 용이한 GMDH 알고리즘을 사용함으로써 전력수요에 영향을 주는 여러 요인들(GDP, 수출, 수입, 취업자 수, 경제활동인구, 석유소비량, 평균기온)을 입력 데이터로 적용하여 분기별로 전력수요량을 예측하였다.

2. 수정된 GMDH 알고리즘

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

$$\begin{aligned} x(k+1) &= \Phi[x(k), u(k)] \\ y(k) &= \Psi[x(k)] \end{aligned} \quad (2.1)$$

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.

$$\begin{aligned} y(k+1) &= f[y(k), y(k-1), \dots, y(k-n+1), \\ &u(k), \dots, u(k-m+1)] \end{aligned} \quad (2.2)$$

이때 $f[\cdot]$ 는 비선형 함수이다.

GMDH는 입력 노드(input node)들의 집합으로 이루어져 있으며, 이 입력 노드들로부터 출력을 얻기 위하여 중간층들을 거치게 된다. 각 층에서의 노드들은 바로 전단계 층의 노드들의 출력들로부터 두개씩을 취하여 출력을 만들어낸다. GMDH의 구성도를 그림2.1에 나타내었다. 그림2.1에서 $z_{s,t}$ 는 s번째 층과 t번째 노드의 출력을 나타내며 m은 입력의 갯수이다. 두번째 층은 i개의 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.

$$\begin{aligned} z_{s,t} &= a_{s,t}z_{(s-1),u}^2 + b_{s,t}z_{(s-1),v}^2 + c_{s,t}z_{(s-1),u} \\ &+ d_{s,t}z_{(s-1),v} + e_{s,t}z_{(s-1),u}z_{(s-1),v} + f_{s,t} \end{aligned} \quad (2.3)$$

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 u번째와

v번째 노드를 가리킨다.

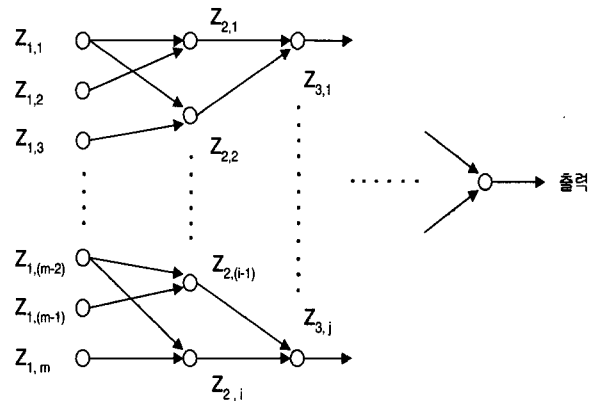


그림 2.1 GMDH의 구조
Fig2.1 Structure of GMDH

다시 말해서, GMDH의 원리는 예를 들어 3번째 층으로 들어오는 i개의 입력들의 조합의 수(C_i^2)에 의해서 다음 층의 노드 수가 결정되고 난 후, 이 노드들의 출력이 식(2.3)에 의해 각각 계산되며, 계산된 출력의 적합성 여부를 성능 기준에 의하여 판별한다는 것이다. 이처럼 각각의 층에서의 적합치 못한 노드들을 판별해내기 위해서는, 다음과 같이 원하는 최종 출력과 실제 노드 출력들 사이의 에러를 계산하는 것이다.

$$r_{s,t} = |y(k) - (z_{s,t})_n| \quad (2.4)$$

여기서 $(z_{s,t})_n$ 은 s번째 층과 t번째 노드의 출력인 $z_{s,t}$ 벡터의 마지막 요소인 n번째 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 알고리즘은 일반적으로 사용되고 있는 신경망 구조와는 달리 중간층의 수와 노드들의 수가 미리 정해지는 것이 아니다.

일반적인 GMDH 알고리즘에서는 데이터를 취사선택하는 기준으로서 미리 정한 일정한 값 R 을 에러 $r_{s,t}$ 와 비교하여 $r_{s,t}$ 가 R 보다 작으면 그 노드의 출력을 취하고 $r_{s,t}$ 가 R 보다 크면 그 노드의 출력은 버린다. 그러나 이러한 방식을 사용하면, 층이 증가함에 따라서 노드의 출력이 실제의 출력에 점점 가까워지기 때문에 잔존하는 노드의 수가 점점 증가한다. 따라서 알고리즘의 계산량이 증가하는 단점이 발생한다. 그러므로 본 논문에서는 기준 R 을 순차적으로 감소 시킴으로써 계산량의 과다를 방지하는 방법을 사용한다.

우선 알고리즘의 종료 방법을 설명하면 각 층에서의 에러의 최소값 IN_s ($s=1, 2, \dots$)를 구하여 이 IN_s 가 최소가 되는 층에서 알고리즘을 종료한다. 이 과정을 그림2.2에 나타내었다. 그림2.2에서 점으로 표시된 것이 IN_s 이다. 그러므로 s 층에서 최소의 에러를 갖는 노드의 출력이 실제의 출력의 추정치가 된다.

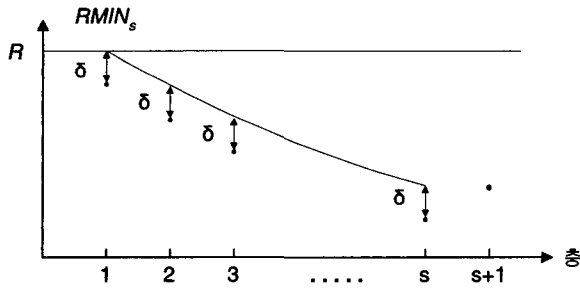


그림 2.2 알고리즘의 종료 방법
Fig 2.2 Finish method of algorithm

그림 2.2에서 알 수 있듯이 일반적인 GMDH 알고리즘에서는 에러가 R보다 작은 노드들이 모두 선택되며 층이 증가함에 따라 IN_s와 R 사이의 간격이 커지므로 노드의 수가 크게 증가한다. 따라서 본 논문에서는 노드의 수가 크게 증가하는 것을 막기 위해 그림 2.2에 표시했듯이 다음과 같은 기준을 선정한다.

$$R_s = IN_s + \delta \quad (2.5)$$

여기서 δ 는 미리 정한 일정한 값이다. 선택 기준을 식(2.5)와 같이 정함으로써 노드의 수가 크게 증가하는 것을 막을 뿐만 아니라 데이터를 IN_s 에 대하여 균등하게 선택할 수 있다는 것을 알 수 있다.

3. 예측 알고리즘의 구현

식(2.2)로 나타낸 비선형 시스템을 모델링하기 위한 GMDH 알고리즘을 다음과 같이 3 단계로 나누어 설명할 수 있다.

단계 1. 변수벡터의 구성

식 (2.3)에 의하여 k 순간에서 s층의 t번째 노드의 출력 $z_{s,t}(k)$ 를 다음 식과 같이 간단하게 표현할 수 있다.

$$z_{s,t}(k) = X_{s-1} \theta_s \quad (3.1)$$

이때

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

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

출 령	입 령							
$y(k-p+2)$	$u(k-p+1)$	$u(k-p)$...	$u(k-m-p+2)$	$y(k-p+1)$	$y(k-p)$...	$y(k-n-p+2)$
$y(k-p+3)$	$u(k-p+2)$	$u(k-p+1)$...	$u(k-m-p+3)$	$y(k-p+2)$	$y(k-p+1)$...	$y(k-n-p+3)$
.								
.								
.								
$y(k)$	$u(k-1)$	$u(k-2)$...	$u(k-m)$	$y(k-1)$	$y(k-2)$...	$y(k-n)$
$y(k+1)$	$u(k)$	$u(k-1)$...	$u(k-m+1)$	$y(k)$	$y(k-1)$...	$y(k-n+1)$

그림 3.1 GMDH의 입력과 출력
Fig 3.1 Input and output of GMDH

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. $(k+1)$ 인 순간에 첫번째 층의 입력과 출력은 그림 3.1과 같다. GMDH로 비선형 다이내믹 시스템을 모델링할 때 인가되는 입력 벡터는 $(m+n)$ 개이다. 각 벡터의 요소는 p개이며 임의로 정해진다. p의 크기에 따라서 인가되는 데이터의 양이 결정된다. 식 (3.1)은 θ_s 와 함께 2차 방정식을 만들어낸다. 이때 $z_{s,t}(k)$ 를 얻기 위해서는, 먼저 식 (3.1)을 풀어 계수 벡터인 연결강도 θ_s 를 얻어야 한다. 각 노드의 출력 $z_{s,t}(k)$ 가 최종적으로 시스템의 출력 $y(k)$ 가 되도록 하는 것이 이 알고리즘의 목적이므로 그 일반식은 다음과 같다.

$$\theta_s = (X_{s-1}^T X_{s-1})^{-1} X_{s-1}^T y(k) \quad (3.2)$$

단계 2. 영향이 적은 데이터의 제거

단계 1에서 구해진 모든 노드들의 출력들이 다음 층으로 전달되는 것은 아니다. 즉 적자 생존의 원칙을 적용해야 한다. 일반적인 GMDH에서는 출력과 새로 생성된 데이터를 훈련 부분과 검사 부분으로 나누어야 한다. 그래서 출력과 생성된 데이터의 훈련부분 과정에서 구해진 평균계급 에러에 의해 필요없는 노드들을 제거한다. 이를 수행하는 방법들에는 여러 가지가 제시되고 있으나 본 논문에서는 비선형 시스템을 동적으로 모델링하고자 하므로 그림 3.1의 모든 데이터를 훈련 부분에서 사용하고 식 (2.4)와 같이 데이터 벡터의 마지막 요소를 검사 부분으로 사용하였다.

$r_{s,t}$ 값의 증가순으로 s층의 모든 $z_{s,t}(k)$ 를 다시 정렬시킨다. 그리고 기준값 R_s 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 IN_s 를 전 단계의 IN_{s-1} 과 비교하는 것이다. 그 중에서 최소값 IN_s 를 구하여 식 (2.5)로 R_s 를 구한다. 그림 2.2와 같이 $IN_s < IN_{s-1}$ 이면 단계 1과 2를 반복하고, $IN_s \geq IN_{s-1}$ 이면 $RMIN_{s-1}$ 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 다항식이 된다.

4. 시뮬레이션 및 결과 분석

본 장에서는 개발된 GMDH 알고리즘을 전력 수요 예측에 적용하였다. 적자 생존 기준값을 결정하는 δ 값은 0.045로 결정하였다.

표 4.1은 사용된 데이터로서 종속변수는 전력수요량(Y), 독립변수는 GDP(X1), 수출(X2), 수입(X3), 취업자(X4), 경제활동인구(X5), 석유소비량(X6), 평균기온(X7)를 사용하였다.

본 장에서는 1999년 1/4분기부터 2001년 1/4분기까지의 전력 수요량의 정확한 예측을 목표로 시뮬레이션하였다.

개발된 알고리즘으로 예측 목표기간으로 정한 1999년 1/4분기부터 2001년 1/4분기까지의 전력 수요량의 정확한 예측을 위한 입력 데이터들의 최적기간을 결정한다. 시뮬레이션

표 4.1 독립변수와 종속변수

Table 4.1 Independent variable and dependent variable

기간(분기)	종속변수	독립변수						
	y	x1	x2	x3	x4	x5	x6	x7
	전력 수요량	GDP	수출	수입	취업자	경제활동인구	석유 소비량	평균 기온
	GWH	조원	억불	억불	천명	천명	천 BBL	°C
1991. 1/4	24625	63.6	152.3	195.7	19085	19486	107942	2.7
1991. 2/4	25387	70	186.5	206	19168	19570	95406	17.5
1991. 3/4	27295	72.3	169.8	201	18920	19338	95279	23.6
1991. 4/4	27069	81.8	210.1	212.6	18166	18703	126037	9.2
1992. 1/4	27641	68.8	169.7	209	19462	19900	137960	4.3
1992. 2/4	28078	74.7	197.5	210	19396	19835	119244	17.0
1992. 3/4	29710	75.2	191.5	193.6	19106	19558	108986	24.1
1992. 4/4	29814	84.8	207.6	205.1	18327	18925	148034	9.0
1993. 1/4	30290	71.7	181.8	199.6	19610	20176	156246	3.4
1993. 2/4	31263	78.4	207.4	213.3	19765	20285	121180	17.1
1993. 3/4	32718	80	203.9	207.8	19610	20132	121421	22.3
1993. 4/4	33524	89.9	229.2	217.2	19130	19730	165723	9.0
1994. 1/4	34206	78	198.8	225.5	20154	20647	168821	3.0
1994. 2/4	35315	84.4	238.3	248.1	20253	20703	136582	18.5
1994. 3/4	39208	86	235.1	257.2	20085	20503	136975	26.2
1994. 4/4	37812	98	287.9	292.7	19648	20136	179120	10.4
1995. 1/4	38441	85.5	261.9	305.2	20637	21043	186857	3.5
1995. 2/4	39661	92.6	320.5	347.4	20777	21177	149911	16.9
1995. 3/4	42839	94.4	320.7	344.4	20667	21054	146108	24.1
1995. 4/4	42330	104.8	347.4	236.1	20015	20471	194346	8.5
1996. 1/4	43682	91.5	315	353.4	20989	21405	204840	2.4
1996. 2/4	44149	98.9	331.8	374.9	21150	21543	157682	17.0
1996. 3/4	47546	100.5	295.8	374.2	21114	21552	155006	24.2
1996. 4/4	47094	111.8	354.7	383.1	20510	21155	203534	9.2
1997. 1/4	47683	96	297.2	370.5	21382	21931	129670	3.5
1997. 2/4	48760	105.1	355.3	373.6	21402	21872	174501	18.1
1997. 3/4	53097	106	341.8	356.9	21132	21692	179042	24.3
1997. 4/4	51243	115.9	367.3	345.2	19762	20940	223691	9.9
1998. 1/4	46837	91.6	322.3	236.6	20244	21725	174949	4.9
1998. 2/4	47122	96.7	348.8	235.4	20049	21646	155773	18.7
1998. 3/4	50042	97.4	304.8	214.6	19924	21511	155192	24.7
1998. 4/4	49471	109	347.2	246.2	19105	20854	184363	10.5
1999. 1/4	49719	96.9	302.5	255.7	20362	21797	195090	4.3
1999. 2/4	51884	107.6	357.4	287.6	20695	21914	161589	18.3
1999. 3/4	56522	110.1	350.9	297.6	20962	21972	170593	24.5
1999. 4/4	56089	123.1	426	356.7	20313	21405	192384	9.3
2000. 1/4	58392	109.2	392.7	388.3	21268	22108	203422	3.0
2000. 2/4	57905	118.1	434.1	398.1	21395	22204	176153	17.5
2000. 3/4	62431	120.3	443.8	404	21266	22082	171287	24.5
2000. 4/4	60803	128.8	452.1	414.4	20403	21432	192772	9.5
2001. 1/4	63576	113.2	401	380.6	21677	22468	200579	3.2

결과들을 분석해 보면 목표로 설정한 분기 수와 동일하게 1996년 4/4분기부터 1998년 4/4분기까지 9개의 분기를 입력으로 한 결과에서 평균에러율이 0.96%로 가장 적게 나타났다. 그림 4.1은 입력 기간에 따른 평균 에러율을 비교한 결과이다.

따라서, 이 연구의 목표 예측 기간인 1999년 1/4분기부터 2001년 1/4분기까지 9개의 분기를 예측하기 위한 최적 입력 기간은 목표로 설정한 분기 수와 동일한 1996년 4/4분기부터 1998년 4/4분기까지임을 알 수 있었다.

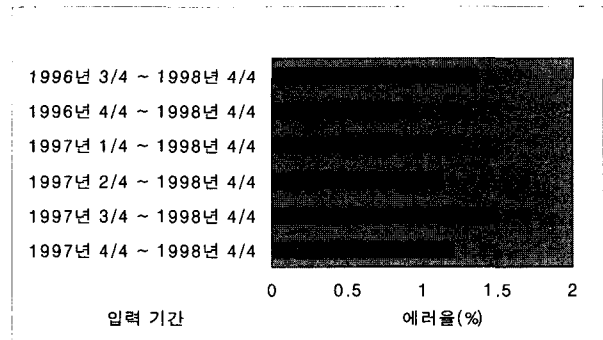


그림 4.1 최적 입력 기간에 대한 시뮬레이션 결과 비교
Fig 4.1 Simulation result comparison for optimum input period

결정된 입력 데이터의 최적 분기 수를 이용하여 입력 데이터와 예측값의 시간적 연관성에 대한 분석을 하였다. 목표 예측 기간인 1999년 1/4분기부터 2001년 1/4분기까지의 전력 수요량의 정확한 예측을 위한 입력 데이터들의 최적 입력 기간이 목표로 설정한 분기 수와 동일한 1996년 4/4분기부터 1998년 4/4분기까지임을 증명하였고, 최적 입력 기간에서 한 분기전의 9개 분기의 데이터를 입력으로 하여 입력 데이터가 과거로 갈수록 예측의 정확도가 어떻게 변하는 가를 시뮬레이션하였다.

시뮬레이션 결과를 보면 목표로 설정한 분기 수와 동일하게 9개의 분기를 입력 기간으로 주었지만 입력 기간이 과거로 갈 수록 평균에러율이 크게 증가함을 볼 수 있다. 그림 4.2는 입력 기간에 따른 평균 에러율 증가 추이를 그래프로 나타낸 것이다.

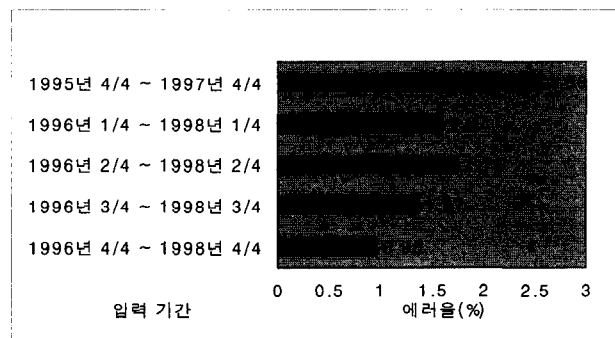
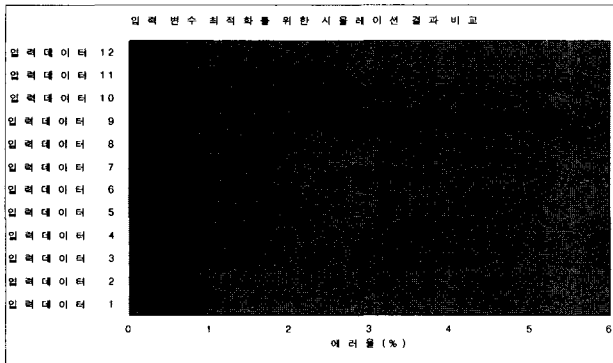


그림 4.2 시간적 연관성에 의한 시뮬레이션 결과 비교
Fig 4.2 Simulation result comparison by time relevance

입력한 독립변수들 중에서 전력 수요량과 연관성이 떨어지는 독립변수들을 제거하여 최적의 입력데이터로 전력 수요를 예측하는 방법을 제시하였다. 본 논문에서는 입력 데이터

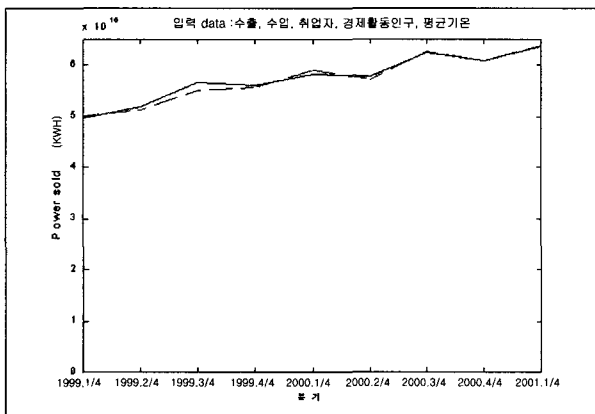
로 GDP, 수출, 수입, 취업자 수, 경제활동인구, 석유소비량, 평균기온이 사용되었다. 이 7개의 데이터중 일부를 입력 데이터로 사용하여 목표 예측 기간인 1999년 1/4분기부터 2001년 1/4분기까지의 전력 수요량을 예측하였다.

그림 4.3의 시뮬레이션 결과들을 비교하여 보면 입력 데이터의 선택이 예측의 정확도에 큰 영향을 미치는 것을 볼 수 있다. 입력 데이터 중에 GDP와 석유소비량은 전력 수요량에 영향을 미치지 못하는 독립 변수임을 분석할 수 있다. 따라서, 전력수요예측 시에 입력 데이터로 사용해야하는 독립변수는 GDP와 석유소비량을 제외한 수출, 수입, 취업자 수, 경제활동인구, 평균기온 임이 증명되었다.



- 입력데이터12 : 수출, 수입, 취업자 수, 경제활동인구, 평균기온(최적 입력 사용) Case 22
- 입력데이터11 : GDP, 수출, 수입, 취업자 수, 경제활동인구 Case 21
- 입력데이터10 : GDP, 수출, 수입, 석유소비량, 평균기온 Case 20
- 입력데이터 9 : 취업자 수, 경제활동인구, 석유소비량, 평균기온 Case 19
- 입력데이터 8 : GDP, 수출, 수입, 취업자 수, 경제활동인구, 석유소비량 Case 18
- 입력데이터 7 : GDP, 수출, 수입, 취업자 수, 경제활동인구, 평균기온 Case 17
- 입력데이터 6 : GDP, 수출, 수입, 취업자 수, 석유소비량, 평균기온 Case 16
- 입력데이터 5 : GDP, 수출, 수입, 경제활동인구, 석유소비량, 평균기온 Case 15
- 입력데이터 4 : GDP, 수출, 취업자 수, 경제활동인구, 석유소비량, 평균기온 Case 14
- 입력데이터 3 : GDP, 수입, 취업자 수, 경제활동인구, 석유소비량, 평균기온 Case 13
- 입력데이터 2 : 수출, 수입, 취업자 수, 경제활동인구, 석유소비량, 평균기온 Case 12
- 입력데이터 1 : GDP, 수출, 수입, 취업자 수, 경제활동인구, 석유소비량, 평균기온 Case 11

그림 4.3 입력 변수 최적화를 위한 시뮬레이션 결과 비교
Fig 4.3 Simulation result comparison for input variable optimization



- 입력 기간 : 1996년 4/4분기 ~ 1998년 4/4분기
- 예측 기간 : 1999년 1/4분기 ~ 2001년 1/4분기
- 입력데이터 : 수출, 수입, 취업자 수, 경제활동인구, 평균기온
- 출력데이터 : 전력수요량(— : 기준값 - - - : 예측값)

그림 4.4 최적 입력 데이터를 사용한 전력 수요 예측 결과
Fig 4.4 Electric power demand forecasting result using optimal input data

결과적으로 최적화된 시뮬레이션 결과는 그림 4.4와 표 4.2와 같이 구할 수 있었다. 1999년 3/4분기 때 최고의 에러율 2.63%를 나타내고, 2000년 4/4분기 때 최저의 에러율 0.15%를 나타내며, 전체적으로 0.96%의 낮은 에러율을 나타낸다. 그래프에서 보면 알 수 있듯이 예측 전력 수요량의 동향이 실제 전력 수요량의 동향과 일치하여 본 논문에서 제안한 예측 알고리즘의 우수성을 나타내고 있다.

표 4.2 출력 데이터와 에러율
Table 4.2 Output data and error rate

분 기	실제 전력 수요량(KWH)	예측 전력 수요량(KWH)	에 러 율 (KWH)	에 러 율 (%)
1999.1/4	49719000000	50069000000	350110000	0.70
1999.2/4	51884000000	51221000000	663040000	1.28
1999.3/4	56522000000	55038000000	1483800000	2.63
1999.4/4	56089000000	55751000000	337930000	0.60
2000.1/4	58392000000	59045000000	652810000	1.12
2000.2/4	57905000000	57341000000	563570000	0.97
2000.3/4	62431000000	62800000000	369240000	0.59
2000.4/4	60803000000	60895000000	91688000	0.15
2001.1/4	63576000000	63936000000	359530000	0.57
계			541300000	0.96

5. 결 론

본 논문에서는 신경회로망을 이용한 기법에 비하여 데이터의 효율적인 활용과 정확성에서 보다 우수한 특성을 보이는 GMDH 알고리즘을 전력수요예측에 적용함으로써 다양한 종류의 데이터를 입력으로 사용하여 정확한 예측을 가능하게 하였다. 또한 본 논문에서 사용한 기법은 연관성이 없는 변수를 제거함으로써 입력 독립변수를 최적화할 수 있기 때문에 전력수요예측에 있어서 입력 변수 선정의 복잡성에 대한 해결 방안을 제시하였다. 그리고 입력 데이터로 경제적인 요인(GDP, 수출, 수입, 취업자 수, 경제활동인구, 석유소비량)과 기후적인 요인(평균기온)을 모두 사용하여서 불규칙한 변동 특성을 갖는 전력수요예측을 더 정밀하게 할 수 있는 방법을 제시하였다.

참 고 문 헌

- [1] 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.
- [2] 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.
- [3] 안기범, "GMDH를 이용한 모델링에 관한 연구," 동아대학교 대학원, 석사 학위 논문, 1995년.
- [4] 이창복, "GMDH 알고리즘을 응용한 전처리 선별과 인공신경망을 이용한 한국종합주가지수 예측에 관한 연구," 연세대학교 대학원, 석사 학위 논문, 1999년.

저 자 소개



이동철(Dong Chul Lee)
 2000년 2월 : 인천대학교 전자공학과 졸업
 2002년 2월 : 인천대학교 대학원
 전자공학과 졸업
 2002년 현재 : 액트통신기술(주)통신연구소
 주임연구원
 관심분야 : GMDH, 유무선 통신

홍연찬(Yeon Chan Hong)

제7권 제 3호 참조

[5] G. Gross and F.D Galiana, "Short-term load forecasting," Proceedings of the IEEE, Vol.75, No.12, pp.1558-1573, 1987.

[6] 이대봉, "직교다항회귀 모델을 이용한 전력 수요 예측," 경남대학교 대학원, 석사 학위 논문, 1997년.

[7] 연형선, "특수일 전력 수요 예측을 위한 신경회로망 시스템의 개발," 강원대학교 대학원, 석사 학위 논문, 1999년.

[8] 허남철, "더미변수를 이용한 최대전력예측," 경남대학교 대학원, 석사 학위논문, 1999년

[9] T. Nishikawa and S. Shimizu, "The characteristics of a based estimator applied to the adaptive GMDH," Mathl. Comput. Modelling, vol. 17, no.1, pp. 37-48, 1993.

[10] S. J. Farlow, Self-Organizing Methods in Modelling : GMDH Type Algorithms. New York : Marcel Decker Inc., 1984.

[11] A. G. Ivakhnenko and A. A. Zholnarskiy, "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.

[12] I. Hayashi and H. Tanaka, "The fuzzy GMDH algorithm by possibility models and its application," Fuzzy Sets and Systems, vol. 36, pp. 245-258, 1990.

[13] P. Y. Chao, P. M. Ferreira, and C. R. Liu, "Applications of GMDH-type modeling in manufacturing," Journal of Manufacturing Systems, vol. 7, no.3, pp. 241-253, 1988.

[14] D. T. PHAM and X. LIU, "Modeling and prediction using GMDH networks of Adalines with nonlinear preprocessor," INT. J. SYSTEMS SCI., vol. 25, no. 11, pp. 1743-1759, Mar. 1994.

[15] Yasuhide Shinohara, Tadashi Dohi and Shunji Osaki, "Predictive Evaluation for Software Testing Progress via GMDH Networks," Electronics and Communications in Japan, Part 3, Vol. 82, No. 5, 1999.

[16] 최종현, 홍연찬, "DPNN을 이용한 비선형 시스템의 모델링에 관한 연구," '95 대한전자공학회 회로 및 시스템 연구회, 전력전자연구회 합동 학술대회, pp. 105-108, 1995년 10월.

[17] 최종현, 홍연찬, "GMDH 알고리즘을 이용한 비선형 시스템의 모델링에 관한 연구," '95 대한전자공학회 추계종합 학술대회, 1995년 12월.

[18] 최종현, "DPNN을 이용한 비선형 시스템의 모델링 및 적응 제어에 관한 연구," 인천대학교 대학원, 석사 학위 논문, 1995년.