

더미변수(Dummy Variable)를 포함하는 다변수 시계열 모델을 이용한 단기부하예측

論 文
52A-8-3

Short-Term Load Forecasting Using Multiple Time-Series Model Including Dummy Variables

李景勳* · 金鎮吾**
(Kyung-Hun Lee · Jin-O Kim)

Abstract - This paper proposes a multiple time-series model with dummy variables for one-hour ahead load forecasting. We used 11 dummy variables that were classified by day characteristics such as day of the week, holiday, and special holiday. Also, model specification and selection of input variables including dummy variables were made by test statistics such as *AIC*(Akaike Information Criterion) and *t-test* statistics of each coefficient. *OLS* (Ordinary Least Squares) method was used for estimation and forecasting. We found out that model specifications for each hour are not identical usually at 30% of optimal significance level, and dummy variables reduce the forecasting error if they are classified properly. The proposed model has much more accurate estimates in forecasting with less *MAPE* (Mean Absolute Percentage Error).

Key Words : short-term load forecasting, dummy variable, multiple time-series model, *AIC*(Akaike Information Criterion), *t-value*

1. 서론

전력부하 예측의 정확도 향상은 발전 비용의 감소와 전력 공급의 신뢰도 향상의 결과로 나타나기 때문에 장기 및 단기 기간의 전력 부하의 예측은 전력 계통의 경제적 운용의 측면에서 매우 중요하다. 특히 그 결과는 전력 계통의 제어 및 전력조류계산의 입력 자료로 사용되기 때문에 지금까지 다양한 방식에 의한 예측 기법이 제안되었다. 이러한 종래의 방법으로는 크게 시계열 방법 [1]과 인공지능 기법의 신경회로망[2,3]이 있다. 이 중에서 시계열 방법, 특히 자기회귀(Autoregression, AR) [4], 통합혼합모형(Autoregressive Integrated Moving Average, ARIMA) 등은 예측 모형을 설정하는데 있어서 자기 변수만을 고려하므로 예측 모형에서는 너무 단순하고 또한 선형모형이므로 좋은 예측 결과를 기대할 수 없다. 이러한 시계열 방법과 비교하여 부하 예측 문제에 내재하는 비선형성, 유연한 입력 선택, 모델링의 복잡성 등과 같은 문제를 적절히 해결할 수 있는 방법으로 신경회로망이 알려져 있으나 기본적으로 몇 가지 단점이 있다. 즉, 학습을 통해 유추된 예측모델의 규칙을 표현할 수 없고, 최적의 성능을 얻기 위한 다층 퍼셉트론(Multilayer Perceptron, MLP)의 은닉층의 수와 각 은닉층의 노드 수를 결정하는 효과적이며 체계적인 방법이 존재하지 않는다 [5].

일반적으로 부하패턴은 평일, 주말, 특수일, 명절 등에 따

라 다르다. 이에 기존의 부하예측 관련 논문은 주말을 제외하거나 [5], 혹은 평일과 주말을 분리하여 부하예측을 행하는 것 [6-7]이 대부분이었다. 그러나 전력부하를 시간에 대해 연속성을 갖는 시계열 자료(Time-Series Data)로 간주하여 예측모델을 설정한다면 부하데이터를 유형별로 분류하여 예측하는 것보다 더 좋은 결과를 기대할 수 있다. 본 논문에서는 예측모델에 반영될 입력변수 선택의 기준으로서 검정통계량(Test Statistics)을 이용하고, 부하패턴별 예측 결과의 향상을 위하여 더미변수(Dummy Variable)를 도입하는 다변수 시계열 예측모델을 제안한다. 또한 사례 연구를 통해 실제 부하 데이터를 입력으로 하여 각 시간에 대한 예측 모형을 수립한 후, 이 모형에 의해 부하를 예측하였으며 그 결과를 뉴로-퍼지 모델 [5]의 결과와 비교하여 그 우수성을 입증하였다.

2. 부하의 특성

기상이나 특수일과 같은 부하의 변동 요인이 없다면 하루의 부하 곡선은 보통 화요일부터 금요일까지 그대로 유지된다. 그림 1은 2000년 12월 4일 월요일부터 12월 10일 일요일까지 한 주간의 부하 특성 곡선을 나타내는데 토요일, 일요일, 월요일은 평일에 비해 전 시간대에 걸쳐 혹은 시간대별로 낮은 수요치를 보이지만 전 주의 동일 요일의 부하 곡선과는 유사한 경향을 보인다. 따라서 부하 예측에 있어서 한 주간의 요일을 보통 화요일부터 금요일까지의 평일, 토요일, 일요일, 월요일, 그리고 특수일로 나누는 것이 일반적이었다. 그러나 전력 부하를 시간에 대해 연속성을 갖는 시계열 자료로 간주하여 사전 예측 모델을 설정한다면, 부하 데

* 準 會 員 : 漢陽大 大學院 電氣工學科 碩士

** 正 會 員 : 漢陽大 電氣工學科 副教授 · 工博

接 受 日 字 : 2002年 12月 11日

最 終 完 了 : 2003年 5月 26日

이더를 유형별로 분류하여 예측하는 것보다 더 좋은 결과를 기대할 수 있다. 따라서 본 논문에서는 요일을 유형별로 분류하는 대신 더미변수(Dummy Variable)를 도입함으로써 부하의 요일별 패턴을 예측모델에 반영하는 시계열 분석을 행하였다.

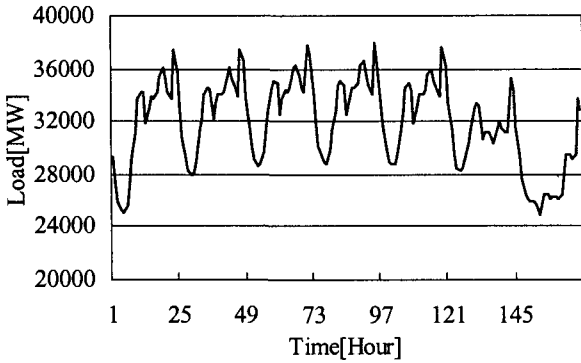


그림 1 한 주간의 부하특성곡선
 Fig. 1 Load profiles of a typical week(Monday, Dec. 4 to Sunday, Dec. 10, 2000)

3. 다변수 시계열 모델(Multiple Time-Series Model)

3.1 임계값(Threshold Value) 선정

더미변수는 특성의 존재 유무에 따라 1 또는 0의 값을 할당함으로써 종속변수의 질적 변화를 나타내기 위한 변수로 쓰인다[8]. 본 논문에서는 표 1과 같은 입력 더미변수의 구성을 통하여 특정일의 더미변수만을 1의 값을 취하여 요일별 부하패턴이 미치는 영향도(影響度)를 종속변수, 즉 예측부하에 반영하고자 한다. 공휴일의 경우, 해당 공휴일이 주말이면 부하패턴은 평상시의 주말과 크게 다르지 않은 반면, 평상일이면 대체로 작게 나타난다. 따라서 공휴일의 경우는 표 1과 같이 주말과 평일을 구분하였다. 표 1의 공휴일과 명절(신정, 설날, 추석)의 더미변수는 다른 더미변수보다 우선권을 갖는다. 예를 들어 토요일이면서 공휴일인 경우, 더미변수 D_7 에만 1의 값이 할당된다. 즉 화요일을 제외한 모든 날은 11개의 더미변수 중 하나의 더미변수에 대하여 1의 값을 갖고 모든 더미변수가 0인 경우는 화요일에 해당된다.

표 1 더미 변수의 구성
 Table 1 Formation of input dummy variables

평 일	공 휴 일	명 절
토요일(D_1)	D_7	전날(D_9) 명절(D_{10}) 다음날(D_{11})
일요일(D_2)		
월요일(D_3)		
수요일(D_4)	D_8	
목요일(D_5)		
금요일(D_6)		

3.2 더미변수(Dummy Variable)를 포함하는 다변수 시계열 모델

다변수 시계열 모델(Multiple Time-Series Model)은 하나의 종속변수와 다변수로 이루어진 독립변수간의 관계를 이용하는 통계적 기법이다. 본 논문에서 제시하는 더미변수의 개념을 도입한 다변수 시계열 모델은 다음 식 (1)과 같다.

$$\hat{L}_h^d = a_0 + \sum_{i=1}^{N_d} b_i L_h^{d-i} + \sum_{j=1}^{N_h} c_j L_{h-j}^d + \sum_{k=1}^{N_D} d_k D_k + \epsilon_t \quad (1)$$

- 여기서, \hat{L} : 부하 추정치(load estimate)
- L : 과거 부하데이터(historical load data)
- D : 더미변수(dummy variable)
- d, h : 예측일, 예측시간(day and hour index of the forecast)
- N_d, N_h : 과거 부하에 대한 일간 및 시간 시차의 개수(daily and hourly lag length)
- N_D : 더미변수의 개수(number of dummy variable)
- $a_0, b_i, c_j,$ and d_k : 각 변수에 대한 회귀계수(regression coefficients for each variable)
- $\epsilon_t \sim iid(0, \sigma^2)$

식 (1)의 오른쪽 식에서 첫째 항은 상수항이고 예측하고자 하는 d 일 t 시각을 기준으로 둘째 항은 과거 N_d 일 동안의 같은 시각의 부하량이며 마찬가지로 셋째 항은 t 시각 바로 직전인 ($t-1$) 시각부터 ($t-N_h$) 시각까지의 부하량을 뜻한다. 또한 넷째 항은 더미변수를 나타내며, 표 1에서 밝힌 바와 같이 본 논문에서 설정한 더미변수의 개수 N_D 는 11이다. 마지막으로 다섯째 항은 평균 0, 분산 σ^2 인 독립적 동일분포를 갖는 확률변수로서 백색잡음(white noise)을 나타낸다. 이 값은 시계열 $\{L_h^{d-i}, L_{h-j}^d, D_k\}$ 와는 독립이라고 가정한다.

식 (1)의 오른쪽 식을 N 개의 관찰치(observation)를 갖는 행렬식으로 다시 표현하면,

$$L_N = X \cdot \beta + \epsilon_t \quad (2)$$

여기서, N : 관찰치의 개수(number of observation)

$$X = [X_1 \ X_2 \ \dots \ X_N]^T$$

$$X_N = [1 \ \sum L_h^{d-i} \ \sum L_{h-j}^d \ \sum D_k]$$

$$\beta = [a_0 \ \sum b_i \ \sum c_j \ \sum d_k]^T: \text{회귀계수}$$

X 는 각 예측시간별로 다르게 설정되는 입력 행렬식으로 행의 길이는 N , 열의 길이는 $(1+N_d+N_h+N_D)$ 보다 작거나 같게 된다. 그 이유는 예측 모형식 (1)의 둘째 항과 셋째 항은 중복되는 경우가 있기 때문이다. 예를 들면 예측

하고자 하는 d 일 t 시각을 기준으로 $(d-1)$ 일의 t 시각은 둘째 항의 L_h^{d-1} , 셋째 항의 L_h^{d-2} 으로 중복 표현될 수 있기 때문이다. 이런 경우 본 논문에서는 둘째 항만을 입력 행렬식 X 에 포함시켰다. 회귀계수 β 의 추정치 $\hat{\beta}$ 는 잔차 제곱합(Residual Sum of Squares, RSS)을 최소로 하는 회귀계수를 구하는 기법인 최소제곱법(Ordinary Least Squares, OLS)을 이용하여 구한다.

4. 입력 변수의 선택과 예측 절차

더미변수를 포함하는 다변수 시계열 모델을 이용하여 부하예측을 수행하기 전에 좀 더 신뢰할 수 있는 예측결과를 얻기 위해서는 적절한 입력변수의 선택이 문제로 남는다. 즉 과거 부하에 대한 일간 및 시간 시차의 개수 N_d, N_h 를 종속변수에 가장 많은 영향을 주는 적절한 값으로 정해야 하는데, 본 논문에서는 다음과 같은 검정통계량(Test Statistics)을 이용하여 입력변수를 선택하였다.

4.1 입력변수의 선택을 위한 검정통계량

(1) AIC(Akaike Information Criterion)

과거의 부하데이터를 이용하여 적절한 일간 및 시간 시차의 개수를 적용함으로써 하나의 종속변수(예측부하)에 대하여 여러 가지 예측모형식을 설정할 수가 있는데 이 중 종속변수를 가장 잘 설명하는 예측모형식을 선택할 필요가 있다. 이때 사용되는 척도가 AIC(Akaike Information Criterion)이다. AIC는 다음과 같은 식 (3)으로 표현되며 더 작은 AIC 값을 갖는 예측모형식이 더 우수하다.

$$AIC = \log \left[\frac{1}{N} \sum_{n=1}^N (L_n - \hat{L}_n)^2 \right] + \frac{2N_k}{N} \quad (3)$$

여기서, N_k : 회귀계수의 개수(number of regression coefficients)

(2) t-값(t-value)

회귀계수에 대한 추정결과는 일반적으로 추정된 회귀계수, 각 회귀계수의 표준오차, 각 회귀계수들에 대한 검정 통계량 t -값을 말한다. 결정계수(Coefficient of Determination, R^2)는 회귀모형식에서 독립변수 전체가 종속변수를 얼마나 설명할 수 있는가를 의미하는데 반하여 t -값은 각각의 독립변수가 종속변수를 얼마나 잘 설명할 수 있는가를 검증하는 통계량이다. 다시 말하면 특정 신뢰수준에 대하여 임계 t -값보다 큰 t -값을 갖는 독립변수일수록 종속변수에 더 많은 영향을 끼친다. 임의의 l 번째 회귀계수에 대한 t -값은 다음 식 (4)를 이용하여 구할 수 있다.

$$t\text{-value}_l = \frac{\hat{\beta}_l}{\sigma \sqrt{d_{ll}}} \quad l \in N_k \quad (4)$$

여기서, $\hat{\beta}_l$: l 번째 회귀계수(l -th regression coefficient)

σ : 표준오차(standard error of regression)

d_{ll} : $(X^T X)^{-1}$ 의 l 번째 대각 성분(l -th diagonal element of $(X^T X)^{-1}$)

이러한 t -값의 신뢰수준별 임계값을 다음 표 2에 정리하였다.

표 2 신뢰수준별 임계 t -값

Table 2 Critical t -values for each significance level

신뢰수준[%]	5	10	20	30	40	50
임계 t -값	1.960	1.645	1.280	1.036	0.842	0.674

4.2 예측 절차

제한한 다변수 시계열 예측모델의 입력값은 과거 부하에 대한 일간 및 시간 시차(lag)를 기반으로 하는 시간별 부하 데이터와 요일 특성에 따른 더미변수로 구성된다. 일간 및 시간 시차의 초기값을 어떻게, 얼마만큼 결정하는가는 예측 결과에 중대한 영향을 미치므로 본 논문에서는 식 (1)에서 더미변수를 표현하는 네 번째 항을 제외한 다음과 같은 식 (5)를 이용하여 초기값으로 결정될 가능성이 있는 일간 및 시간 시차의 값의 조합을 12가지 case로 구성하여 회귀분석하였다.

$$\hat{L}_h^d = a_0 + \sum_{i=1}^{N_d} b_i L_h^{d-i} + \sum_{j=1}^{N_h} c_j L_{h-j}^d + \epsilon_t \quad (5)$$

뉴로퍼지 모델 [5]과의 예측결과 비교를 위하여 본 논문에서 정한 예측구간(1994년 1월 18일부터 4월 30일까지)에 대한 12가지 case의 일간 및 시간 시차의 조합별 회귀분석 결과를 다음 표 3에 나타내었다. 이 표에서 MAPE는 절대 평균 백분율 오차(Mean Absolute Percentage Error, MAPE), Std.는 표준편차(Standard Deviation), Max.는 예측 오차의 최대값을 나타내는데 일간 시차 N_d 는 7, 시간 시차 N_h 는 36일 때 MAPE와 Std.가 가장 작은 값을 가지므로 본 논문에서는 일간 및 시간 시차의 초기값을 각각 7과 36으로 정하였다.

본 논문에서는 그림 2에 나타낸 예측 절차와 같이 두 가지 방법으로 시간별 부하예측을 수행하였다.

Method 1: 일간 시차 N_d 와 시간 시차 N_h 의 초기값의 모든 경우의 수($N_d \times N_h$ 개)만큼 예측모형식을 수립하여 가장 작은 AIC 값을 갖는 일간 및 시간 시차의 값과, 임계 t -값을 정하기 위한 임의의 신뢰수준을 입력한다. 식 (9)와 최소제곱법을 이용하여 회귀계수를 추정한 후 각 독립변수에 대한 t -값을 구한다. 임의의 신뢰수준을 기준으로 자신의 t -값이 임계 t -값보다 큰 독립변수만을 선택하고, 모든 독립변수의 t -값이 임계 t -값보다 큰 값을 가질 때까지 반복한다. 임계 t -값 보다 큰 값을 갖는 독립변수만을 최종 예측모델로 간주하여 회귀분석한 후, 부하를 예측한다.

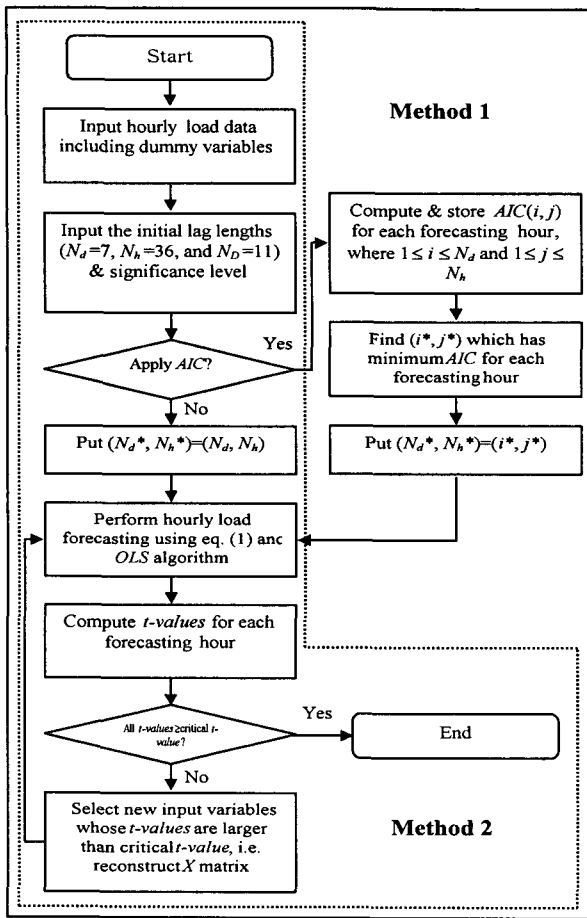


그림 2 부하예측 순서도(Method 1, Method 2)
 Fig. 2 Flowchart of hourly load forecasting (Method 1 & Method 2)

표 3 시차에 대한 회귀분석 결과 (1994.1.18-4.30)
 Table 3 Regression results for potential initial lag lengths (Jan. 18 to April 30, 1994)

Case	N_d	N_h	MAPE	Std.	Max.
1	1	12	0.866	0.621	3.244
2	1	24	0.774	0.581	3.019
3	1	36	0.646	0.533	3.054
4	3	12	0.857	0.624	3.301
5	3	24	0.772	0.588	3.186
6	3	36	0.642	0.535	3.153
7	7	12	0.852	0.615	3.274
8	7	24	0.785	0.584	3.201
9	7	36	0.636	0.534	3.118
10	10	12	0.850	0.607	3.045
11	10	24	0.801	0.577	3.064
12	10	36	0.641	0.524	2.975

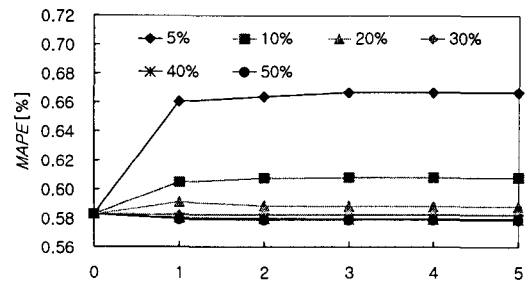
Method 2 : Method 1에서 일간 및 시간 시차의 입력값을 결정하는데 이용한 검정통계량 AIC 를 이용하지 않고 일간 및 시간 시차의 입력값을 각각 초기값 7, 36으로 고정시킨다. 나머지 과정은 Method 1과 동일하다.

5. 사례 연구

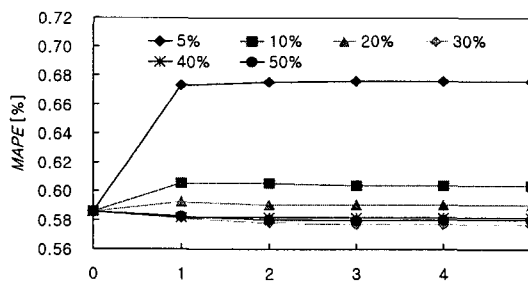
1993년 1월 1일부터 1994년 1월 17일까지 382일 동안의 실제 우리나라의 과거 부하 데이터를 초기 입력으로 하여 Method 1과 2를 통해 각 시간에 대한 예측모델의 모형을 수립한 후, 뉴로-퍼지 모델과의 비교를 위해 참고문헌 [5]에서 제시한 모델의 테스트 범위와 동일한 1994년 1월 18일부터 4월 30일까지 103일 동안의 부하를 시간별로 예측하였다.

5.1 예측결과 비교(Method 1 vs. Method 2)

그림 3은 Method 1과 2의 신뢰수준별 예측결과를 나타낸다. 각 시간별로 최소 AIC 를 갖는 경우의 일간 및 시간 시차의 값을 입력값으로 설정한 Method 1의 경우 신뢰수준 40 [%]에서 가장 작은 $MAPE$ (0.579 [%])값을 갖으며, 일간 및 시간 시차의 입력값을 각각 7과 36으로 고정시킨 Method 2의 경우는 신뢰수준 30 [%]에서 가장 작은 $MAPE$ (0.577 [%])값을 갖는다. 그림 4는 특수일이 포함되지 않은 전형적인 한 주일의 예측결과를 나타낸 그림으로 실제부하와 예측부하가 거의 일치함을 볼 수 있으며, 이 때 $MAPE$ 의 값은 0.6 [%] 이하이다.



(a) Method 1의 결과



(b) Method 2의 결과

그림 3 부하예측 결과 (1994.1.18-4.30)
 Fig. 3 Load forecasting results (Jan. 18 to April 30, 1994)

표 4는 각 시간별 예측결과를 정리한 것으로 Method 1과 2는 각각 신뢰수준 40 [%]와 30 [%]에 대한 결과이며, 5번째 반복 후의 결과이다. 표의 6번째 열은 시간별로 가장 작은 AIC 값을 갖는 일간 및 시간 시차의 입력값이다. 이 표에서 모든 시간의 예측결과가 뉴로-퍼지 모델의 결과와 비

교하이 0.45 [%] 정도의 우수한 예측결과의 차이를 보였으며 시간에 대한 총평균 *MAPE* 는 약 0.42 [%] 정도 낮은 0.579 [%]와 0.577 [%]로 만족할 만한 결과를 얻을 수 있었다. 또한 근소한 차이이긴 하지만 Method 2는 Method 1에 비해 더 좋은 결과를 나타내었다.

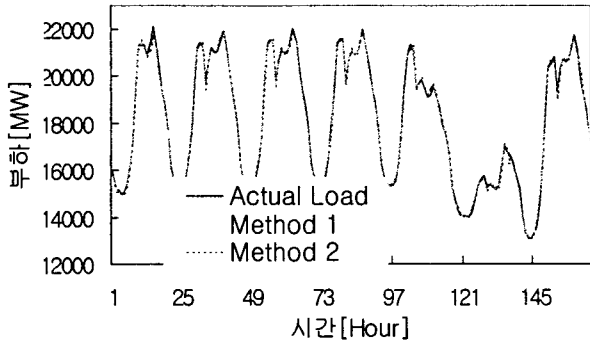


그림 4 한 주간의 부하예측 결과 (1994.1.18-4.30)
 Fig. 4 Load forecasting results for a typical week (Jan. 18 to April 30, 1994)

표 4 시간별 예측결과(MAPE) 비교[%] (1994.1.18-4.30)
 Table 4 Comparison of MAPE (Jan. 18 to April 30, 1994)

시간	뉴로-퍼지모델[5]		다변수 시계열 모델				
			Method 1			Method 2	
	MAPE	Max.	MAPE	Max.	N_d^*, N_h^*	MAPE	Max.
1	1.14	5.21	0.614	3.030	7,32	0.569	2.805
2	0.76	3.08	0.400	1.235	1,29	0.453	1.245
3	0.64	3.20	0.395	1.312	3,32	0.385	1.555
4	0.60	3.52	0.363	1.360	3,19	0.347	1.437
5	0.66	3.30	0.382	1.296	3,26	0.361	1.520
6	0.75	4.69	0.563	1.798	2,26	0.538	1.864
7	1.12	5.83	0.580	2.624	1,22	0.587	2.722
8	1.39	7.05	0.727	2.806	1,33	0.796	2.514
9	1.87	8.84	1.006	5.963	4,29	1.060	4.976
10	0.98	11.87	0.623	3.392	1,25	0.636	2.377
11	0.84	11.28	0.515	2.997	2,14	0.523	2.921
12	0.73	9.33	0.503	1.914	2,32	0.482	1.808
13	0.88	6.61	0.577	3.512	1,25	0.551	3.112
14	0.80	7.61	0.478	2.325	1,33	0.532	2.513
15	0.64	8.58	0.477	3.330	1,17	0.468	3.477
16	0.75	8.04	0.470	3.134	1,26	0.436	2.649
17	0.76	8.58	0.510	2.615	7,27	0.519	2.758
18	1.58	5.79	0.865	3.321	2,36	0.855	3.305
19	1.07	4.83	0.732	3.983	2,36	0.732	3.209
20	2.25	7.67	0.837	3.079	4,35	0.820	3.035
21	0.71	4.43	0.677	2.239	2,31	0.703	2.385
22	0.92	5.35	0.449	1.519	1,36	0.479	1.478
23	0.92	5.93	0.636	1.932	3,26	0.520	1.619
24	1.11	4.94	0.529	2.496	1,29	0.492	2.097
평균	1.00	6.49	0.579	2.634	-	0.577	2.474

5.2 Method 2의 예측결과

이 절에서는 일간 및 시간의 시차를 검정통계량 *AIC*를 이용하여 입력한 Method 1에 비해 더 낮은 *MAPE* 를 보이는 Method 2의 예측결과를 정리하였으며 1994년뿐 만 아니라 2000년 1월 1일부터 4월 30일까지 121일 동안도 함께 예측을 수행하였다. 표 5와 6은 신뢰수준 30 [%]에 대하여 요일별 예측결과를 1994년과 2000년으로 구분하여 정리한 것이다. 이 두 표에서 주말을 포함한 평상일의 예측결과는 만족할만한 수준으로 얻어졌지만, 특수일은 상대적으로 *MAPE* 값이 높게 나타났다. 특히 명절의 경우는 1994년에 1.184 [%], 2000년에 1.283 [%]로 *MAPE* 값이 높게 나타나는데 이는 명절에 해당하는 더미변수의 관찰치가 각각 4개와 3개로서 적은 관찰치를 입력으로 회귀분석 하였으므로 명절에 대한 더미변수의 영향력을 종속변수에 정확히 반영하지 못한 결과로 사료된다. 관찰치의 구간을 더 길게 한다면 명절에 대한 더 좋은 부하예측 결과를 얻을 수 있을 것으로 사료된다.

표 5 요일별 예측결과(MAPE) 비교[%] (신뢰수준: 30[%], 1994.1.18-4.30)

Table 5 Comparison of MAPE[%] for day types (significance level: 30[%], Jan. 18 to April 30, 1994)

요일(날수)	iter. 0	iter. 1	iter. 2	iter. 3	iter. 4	iter. 5
월(14)	0.553	0.539	0.538	0.538	0.538	0.538
화(13)	0.611	0.609	0.607	0.605	0.604	0.604
수(15)	0.543	0.553	0.547	0.546	0.546	0.546
목(14)	0.510	0.495	0.495	0.494	0.494	0.494
금(15)	0.550	0.569	0.565	0.563	0.563	0.563
토(15)	0.629	0.608	0.600	0.599	0.599	0.599
일(14)	0.621	0.622	0.613	0.612	0.612	0.612
특수일(2)	0.839	0.822	0.853	0.854	0.854	0.854
명절(1)	1.356	1.184	1.222	1.226	1.225	1.225
평균	0.586	0.581	0.578	0.577	0.577	0.577

표 6 요일별 예측결과(MAPE) 비교[%] (신뢰수준:30 [%], 2000.1.1-4.30)

Table 6 Comparison of MAPE[%] for day types (significance level: 30[%], Jan. 1 to April 30, 2000)

요일(날수)	iter. 0	iter. 1	iter. 2	iter. 3	iter. 4	iter. 5
월(17)	0.651	0.646	0.647	0.647	0.648	0.648
화(17)	0.571	0.535	0.537	0.538	0.538	0.539
수(15)	0.565	0.537	0.537	0.538	0.539	0.539
목(17)	0.672	0.633	0.630	0.631	0.632	0.632
금(17)	0.584	0.595	0.594	0.595	0.595	0.595
토(16)	0.558	0.524	0.525	0.527	0.527	0.528
일(18)	0.682	0.668	0.667	0.668	0.670	0.670
특수일(2)	0.833	0.802	0.837	0.833	0.835	0.836
명절(2)	1.287	1.283	1.302	1.303	1.306	1.308
평균	0.628	0.608	0.609	0.610	0.611	0.611

표 7과 8은 신뢰수준별 임계 t -값을 기준으로 24시간에 대하여 가장 많이 선택된 입력변수 상위 5개를 1994년과 2000년으로 구분하여 정리한 것이다. 선택빈도는 팔호에 나타내었다. 이 두 표에서 $h-1$ 로 표현한 예측시간의 한 시간 전 부하는 전 시간에 걸쳐 선택되었으며 종속변수에 가장 많은 영향을 끼친다. 두 번째로 많이 선택된 독립변수는 1994년의 경우는 $h-2$ 로 표현한 예측시간의 두 시간 전 부하, 2000년의 경우는 명절 전날 더미변수 D_9 로 나타나는데 이는 종속변수에 영향을 주는 독립변수는 시시각각 변하기 때문이다.

표 7 선택된 독립변수와 선택빈도 (5번째 반복, 1994. 1. 18-4.30)

Table 7 Selected variables and its frequencies (at 5th iteration, Jan. 18 to April 30, 1994)

신뢰수준 [%]	선택빈도 순위				
	1	2	3	4	5
5	$h-1$ (24)	$h-2$ (13)	D_2 (13)	$d-1$ (9)	D_7 (9)
10	$h-1$ (24)	$h-2$ (14)	D_2 (14)	D_7 (12)	D_1 (11)
20	$h-1$ (24)	$h-2$ (18)	D_2 (18)	D_7 (16)	D_1 (15)
30	$h-1$ (24)	$h-2$ (21)	D_2 (18)	D_7 (16)	D_{10} (16)
40	$h-1$ (24)	$h-2$ (21)	D_2 (18)	D_7 (18)	D_{10} (18)
50	$h-1$ (24)	$h-2$ (21)	D_{10} (21)	$h-17$ (20)	D_2 (19)

표 8 선택된 독립변수와 선택빈도 (5번째 반복, 2000.1.1-4.30)

Table 8 Selected variables and its frequencies (at 5th iteration, Jan. 1 to April 30, 2000)

신뢰수준 [%]	선택빈도 순위				
	1	2	3	4	5
5	$h-1$ (24)	D_9 (14)	D_2 (11)	a_0 (9)	$d-1$ (9)
10	$h-1$ (24)	D_9 (15)	a_0 (12)	D_2 (12)	$h-2$ (10)
20	$h-1$ (24)	D_9 (17)	a_0 (16)	D_1 (15)	D_2 (14)
30	$h-1$ (24)	D_9 (18)	a_0 (18)	D_1 (17)	D_2 (15)
40	$h-1$ (24)	D_9 (19)	a_0 (18)	D_1 (17)	D_2 (17)
50	$h-1$ (24)	D_9 (21)	a_0 (19)	$h-25$ (19)	$h-22$ (18)

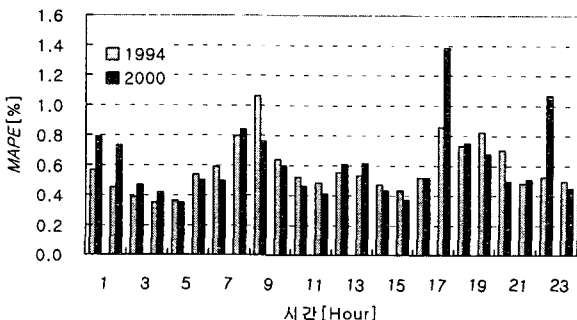


그림 5 1994년과 2000년의 MAPE 비교 (Method 2)
Fig. 5 Comparison of MAPE between the year 1994 and 2000 by Method 2

신뢰수준 30 [%]에 대한 Method 2의 1994년과 2000년의 시간별 예측결과를 그림 5에 나타내었다. 두 예측기간 모두에서 출퇴근 시간인 7시와 10시 사이, 18시와 20시 사이에 상대적으로 MAPE 값은 높게 나타났다. 또한 한밤중인 23시와 4시 사이는 1994년에 비하여 2000년에 상대적으로 높게 나타나는데 이는 에너지 소비패턴의 변화와 정부의 심야 전력 사용 권장제 기인한다고 사료된다.

6. 결 론

본 논문은 더미변수(Dummy Variable)를 이용한 다변수 시계열 모델을 시전예측(One-Hour Ahead Forecasting)에 적용하여 그 결과를 기존의 뉴로-퍼지 모델과 비교하였다. 더미변수는 요일별로 서로 다른 부하패턴의 영향을 종속변수에 반영하는 기능을 한다. 제안한 예측모델은 AIC나 t -값과 같은 적절한 검정통계량을 이용하여 입력 변수 선택의 자동화를 꾀하였으며, AIC의 사용여부에 따라 두 가지 방법으로 시간별 예측모델을 설계하였다. 두 가지 방법 모두에서 제안한 예측모델은 뉴로-퍼지 모델에 비해 우수한 결과를 나타내었으며, AIC를 이용하지 않고 신뢰수준별 임계 t -값을 기준으로 해당 독립변수만을 취하는 방법인 Method 2를 이용하여 더 좋은 결과를 얻을 수 있었다. 또한 각 시간별 부하예측에 영향을 미치는 독립변수는 시간별로 서로 다르며, 요일별 부하패턴을 가장 잘 표현할 수 있는 더미변수의 구성과 일간 및 시간 시차(daily and hourly lag lengths)의 적절한 입력값은 예측결과에 많은 영향을 끼친다는 사실을 확인할 수 있었다. 향후 예측모델에 온도나 습도, 혹은 연간 부하성장률 등을 반영한다면 더 좋은 예측 결과를 기대할 수 있을 것으로 사료된다.

참 고 문 헌

- [1] M. T. Hagan and S. M. Behr, "The Time Series Approach to Short-Term Load Forecasting", IEEE Trans. on Power Systems, Vol. 2, No. 3, pp. 785-791, 1987.
- [2] C. N. Lu and H. T. Wu, "Neural Network Based Short Term Load Forecasting", IEEE Trans. on Power Systems, Vol. 8, No. 1, pp. 336-342, 1993.
- [3] M. Daneshdoost, M. Lotfalian and G. Bumroonggit, "Neural Network with Fuzzy Set-Based Classification for Short-Term Load Forecasting", IEEE Trans. on Power Systems, Vol. 13, No. 4, pp. 1386-1391, 1998.
- [4] A. D. Papalexopoulos and T. C. Hesterberg, "A Regression-Based Approach to Short-Term System Load Forecasting", IEEE Trans. on Power Systems, Vol. 5, No. 4, pp. 1535-1547, 1990.
- [5] 박영진, 심현정, 왕보현, "뉴로-퍼지 모델을 이용한 단기 전력 수요 예측시스템", 전기학회 논문지, 49권 3호, pp.107-117, 2000.
- [6] K. Y. Lee, Y. T. Cha, and J. H. Park, "Short-Term Load Forecasting using an Artificial Neural Network",

IEEE Trans. on Power Systems, Vol. 7, No. 1, pp. 124-132, 1992.

- [7] Alex D. Papalexopoulos, Shangyou Hao, and Tie-Mao Peng, "An Implementation of A Neural Network Based Load Forecasting Model for The EMS", IEEE Trans. on Power Systems, Vol. 9, No. 4, pp. 1956-1962, 1994.
- [8] Joseph F. Hair, Rolph E. Anderson, and Ronald L. Tatham, William C. Black, Multivariate Data Analysis, Prentice Hall, Inc., New Jersey, 1998.

저 자 소 개



이 경 훈 (李 景 勳)

1975년 1월 2일 생. 2000년 한양대 전기공학과 졸업. 2002년 동 대학원 전기공학과 졸업(석사). 현재 LG산전 연구원.

Tel : 02-2290-0347, Fax : 02-2297-1569

E-mail : ceilrule@ihanyang.ac.kr



김 진 오 (金 鎮 吾)

1956년 1월 17일 생. 1980년 서울대 전기공학과 졸업. 1983년 동 대학원 전기공학과 졸업(석사). 1991년 Texas A&M Univ. 전기공학과 졸업(공학박). 현재 한양대 공과대학 전자전기컴퓨터공학부 부교수

Tel : 02-2290-0347, Fax : 02-2297-1569

E-mail : jokim@hanyang.ac.kr