

# 특수일 최대 전력 수요 예측을 위한 결정계수를 사용한 데이터 마이닝

論 文

58-1-4

## Data Mining Technique Using the Coefficient of Determination in Holiday Load Forecasting

魏永民\*·宋敬彬\*\*·朱成官†

(Young-Min Wi · Kyung-Bin Song · Sung-Kwan Joo)

**Abstract** - Short-term load forecasting (STLF) is an important task in power system planning and operation. Its accuracy affects the reliability and economic operation of power systems. STLF is to be classified into load forecasting for weekdays, weekends, and holidays. Due to the limited historical data available, it is more difficult to accurately forecast load for holidays than to forecast load for weekdays and weekends. It has been recognized that the forecasting errors for holidays are large compared with those for weekdays in Korea. This paper presents a polynomial regression with data mining technique to forecast load for holidays. In statistics, a polynomial is widely used in situations where the response is curvilinear, because even complex nonlinear relationships can be adequately modeled by polynomials over a reasonably small range of the dependent variables. In the paper, the coefficient of determination is proposed as a selection criterion for screening weekday data used in holiday load forecasting. A numerical example is presented to validate the effectiveness of the proposed holiday load forecasting method.

**Key Words** : Load forecasting, Polynomial regression, Coefficient of determination

### 1. 서 론

전력 수요 예측은 전력계통의 계획, 유지보수 및 투자, 운용과 전력 시장운영에 필수 도구이다. 또한 최근 진행 중인 전력 시스템 및 시장 관련 다양한 연구, 예를 들어 장기 수급 계획, 분산전원, 수요 반응 등에서 기초 자료로 사용되어지고 있다. 하지만 전력 수요는 그 특성이 비선형적 요소가 강하기 때문에 미래의 전력 수요 예측을 하는 것에 큰 어려움이 있다. 전력 수요가 이러한 비선형적인 특성을 갖는 원인은 전력 수요가 사회·경제적 요소, 시계절적 요소, 기후적 요소, 전력시장 요소, 기타 요소에 의해 영향을 받기 때문이다. 현실적으로 이러한 모든 요소를 감안하여 미래의 전력 수요를 예측하기 위한 시스템을 개발하기는 매우 어려운 일이다.

전력 수요 예측 기법은 매우 다양하게 발전되어 왔다. 초창기 시계열법[1], 회귀분석법 등이 연구 되었으며, 국·내외로 전문가 지식 기반의 인공지능기법[2]-[5]과 퍼지 방법[6]-[7] 통한 예측 방법이 연구되어지고 있다. 또한 최근에는 기존 방법들을 결합한 하이브리드 방식 알고리즘에 대한 연구가 활발히 진행 중이다[8]. 현재까지 진행된 연구 내용에서 알 수 있듯 평일 대상의 전력 수요 예측은 대부분 높

은 정확성을 보인다. 하지만 특수일에 대한 정확도는 평일에 비해 많이 떨어지는 것을 확인 할 수 있다. 국내 연구진에 의해 몇몇 특수일 전력 수요 예측 관한 논문이 국·내외 논문으로 발간되었다. 하지만 평일 수준의 정확성에는 아직 미치지 못하는 것이 현실이다. 따라서 본 논문에서는 이러한 점을 극복하기 위한 결정계수를 이용한 데이터 마이닝과 데이터 마이닝을 통해 얻어진 데이터를 사용하여 특수일 전력 수요 예측을 하는 다항회귀분석 알고리즘을 제시한다.

정확한 전력 수요 예측을 위해서는 비선형적인 요소를 얼마나 정확히 제거하여, 전력 수요 특성을 잘 나타내는 양질의 데이터를 예측 알고리즘에 사용하느냐가 중요한 점으로 작용된다. 본 논문에서 양질의 데이터 즉, 과거 전력 수요의 패턴을 잘 표현하는 데이터를 얻기 위해 통계학에서 사용하는 결정계수를 이용한 데이터 마이닝을 사용하였다. 또한 결정계수를 이용한 데이터 마이닝을 통해 얻어진 데이터를 다항회귀분석의 입력변수로 사용하여 기존 방식인 선형 퍼지 선형회귀보다 정확한 특수일 전력 수요 예측 결과를 얻을 수 있었다.

본 논문에서는 특수일 전력 수요 예측의 정확성을 평일 수준으로 끌어올리기 위한 알고리즘을 제안하고 제시된 방법에 따른 모의 결과를 제시하였다. 2절에서는 온도효과를 이용하여 과거 데이터 선택을 위한 결정계수를 사용한 데이터 마이닝 알고리즘을 제시하며, 데이터 마이닝을 통해 선택된 데이터를 사용한 전력 수요 예측을 위한 다항회귀분석법을 소개한다. 또한 3절에서는 제안된 알고리즘을 이용하여 2006년 특수일 전력 수요 예측을 통하여 제안된 알고리즘을 기존 퍼지 선형회귀분석을 이용한 전력 수요 예측 알고리즘과 비교 검증하였다.

\* 正 會 員 : 高麗大學教 電氣電子電波學科 碩士課程

\*\* 正 會 員 : 崇實大學教 電氣工學府博 副教授·工博

† 교신저자, 正會員 : 高麗大學教 電氣電子電波工學部 助教授·工博

E-mail : skjoo@korea.ac.kr

接受日字 : 2008年 8月 20日

最終完了 : 2008年 9月 22日

## 2. 본 론

### 2.1 전력 수요 예측

과거 국내 전력 수요 예측 관련 연구에서 살펴보면 특수일과 평일(평일, 주말)의 전력 수요 예측 오차의 편차가 큼을 쉽게 확인할 수 있다. 이러한 현상의 가장 큰 이유는 평일 대비 특수일이 가진 과거 데이터의 회소성이 근본적인 원인이다. 정확한 예측을 하기 위해서는 많은 데이터가 있어야 한다. 하지만 데이터가 많아도 양질의 데이터가 부족하다면 절대적인 과거 데이터의 양은 무의미하다. 따라서 전력 수요 예측의 정확성을 높이기 위해서는 과거의 데이터 중에 과거 전력 수요 패턴을 잘 보여주는 양질의 데이터를 얻기 위한 알고리즘이 필요하다. 본 논문은 이를 위해 데이터 간 상관관계를 나타내는 결정계수를 이용한 데이터 마이닝을 통해 과거의 전력 수요 데이터의 온도 영향성을 분석하여 시뮬레이션을 위한 과거 데이터 집합을 구성하였다. 또한 이렇게 구성된 과거 데이터 집합을 이용하여 다항회귀분석을 통해 특수일 전력 수요 예측을 진행하였다.

### 2.2 결정계수를 사용한 특수일 전력 수요 데이터 마이닝

정확한 전력 수요 예측 모델을 위해서는 과거 전력 수요 패턴을 잘 보여주는 과거 데이터가 필요하다. 본 논문에서 제안하는 알고리즘은 특수일과 평일 전력 수요 데이터의 관계를 이용하여 미래 특수일 수요를 예측한다. 정확한 특수일 예측을 위해 과거 특수일이 가진 전력 수요의 특성을 잘 나타내는 과거의 데이터 선택이 필수적이다. 제안된 알고리즘은 온도를 이용하여 특수일과 평일 전력 수요 데이터의 관계의 유사도를 판단하여 평일 과거 데이터를 찾아낸다. 유사도를 측정하기 위해 통계학에서 사용되는 결정계수를 이용한다[9]. 결정계수는 두 변수 간의 설명력이 최대일 때 1이 되고, 최소일 때 0이 된다. 따라서 결정계수는 0과 1 사이에서 존재한다.

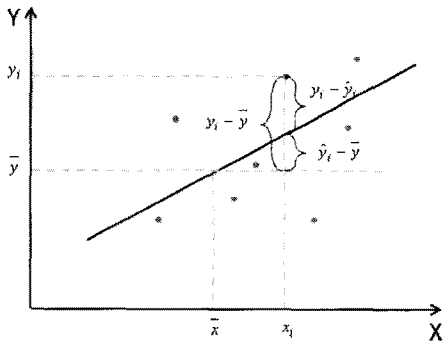


그림 1 결정계수  
Fig. 1 Coefficient of Determination

그림 1과 같이  $x$ 와  $y$  두 변수에 대한 산포도를 보일 경우 결정계수는 수식 (1)-(4)에 의해서 구해진다. 독립변수  $x$ 와 종속변수  $y$ 로 이루어진  $n$ 개의 회귀분석 샘플이 존재할 때, 결정계수는

$$\text{결정계수}(R^2) = \frac{\text{회귀변동}(SSR)}{\text{전체변동}(SST)} \quad (1)$$

$$\text{회귀변동}(SSR) = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2)$$

$$\text{오차변동}(SSE) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

$$\text{전체변동}(SST) = \text{오차변동}(SSE) + \text{회귀변동}(SSR) \quad (4)$$

여기서  $y_i$ 는  $i$ 번째 종속 변수,  $\bar{y}$ 는  $y_i$ 의 평균이며  $\hat{y}_i$ 는  $y_i$ 의 선형회귀 값이다. 결정계수(Coefficient of Determination)는 회귀변동(SSR: Sum of Squared Residuals)과 전체변동(SST: Sum of Squares Total)의 비로 나타낸다. 전체변동은 오차변동(SSE: Sum of Squared Error)과 회귀변동의 합을 의미한다.

본 논문에서는 특수일 전력 수요 예측을 위해 다항회귀분석을 사용한다. 회귀분석의 독립변수  $x$ 로는 특수일 이전의 평일 전력 수요 데이터의 평균을 사용하며, 종속변수로는 특수일 전력 수요 데이터를 사용한다. 과거 평일 데이터 중  $x$ 를 선택하기 위해 본 논문에서는 특수일 최고 기온과 유사한 온도를 갖는 평일 데이터를 추출한다. 제안된 알고리즘은 평일 전력 수요 데이터 선택을 위해 결정계수를 사용하며, 다음에서 소개되는 절차에 의해 추출된 평일 데이터 중에 특수일 온도의 전력 수요 패턴을 보여주는 데이터를 선택하게 된다. 데이터 선정을 위한 절차는 다음과 같다.

- Step 1) 평일 데이터를 온도와 전력 수요에 관해서 각각 정규화 한다.
- Step 2) 정규화된 평일 데이터의 온도와 전력 수요 간의 결정계수와 데이터별 선형회귀의 잔차(Residuals)를 구한다.
- Step 3) 결정계수가 임계값보다 작을 경우, Step 2에서 구한 잔차가 큰 데이터를 데이터 집합에서 삭제한다. 여기서 삭제 대상 데이터는 잔차의  $1-\sigma$  밖의 데이터이다. 만약 결정계수가 임계값보다 큰 경우 Step 6으로 이동한다.
- Step 4) Step 3에서 남은 데이터를 가지고 다시 결정계수와 잔차를 계산한다.
- Step 5) Step 3과 Step 4를 데이터가 4개 남을 때까지 반복한다.
- Step 6) 남은 데이터 중 잔차가 작은 상위 4개의 데이터를 선택한다.

그림 2는 평일 데이터 선택 과정을 나타낸 절차도이며, 그림 3은 결정계수를 이용한 데이터 마이닝을 보여주는 그림이다. 그래프에서  $x$ ,  $y$ 축은 각각 정규화된 온도와 전력 수요를 나타낸다. 그림 2에서 제시한 평일 데이터 선정 절차에 따라 루프연산을 반복할수록 데이터 선정 기준인 결정계수의 값이 1에 가까워짐을 확인할 수 있다.

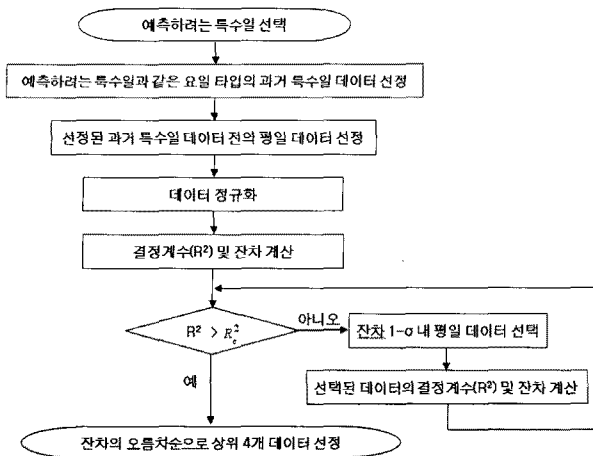


그림 2 평일 데이터 선정 절차도  
Fig. 2 The Flowchart of Selecting Weekday Data

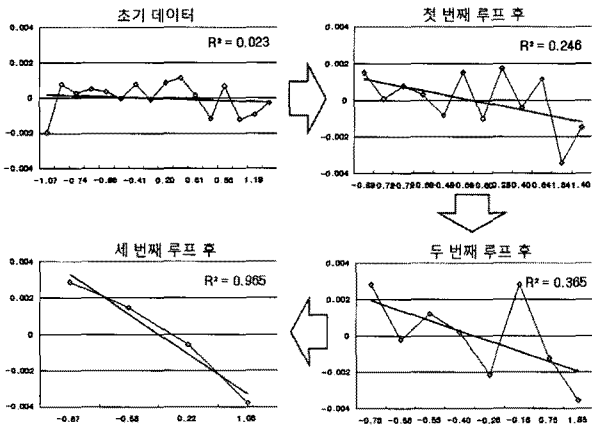


그림 3 결정계수를 이용한 데이터 마이닝  
Fig. 3 Example of Data Mining.

2.3 다항회귀분석

회귀분석이란 어떤 변수(독립변수)를 지정해서 그 지정변수에 대응해서 변화하는 다른 변수(종속변수)의 값을 추정하는 방법이다. 통계학에서는 비선형적인 요소가 강한 데이터에서는 단순회귀분석 보다 다항회귀분석을 주로 이용한다. 왜냐하면 다항회귀분석이 비선형적 요소를 선형회귀보다 더 잘 표현해 줄 수 있기 때문이다. 일반적인 2차 다항회귀 모델은 다음의 수식(4)와 같다.

$$y = \alpha + \beta x + \gamma x^2 \quad (4)$$

여기서  $\alpha, \beta, \gamma$ 는 2차 회귀분석 모델 계수이다.

2차 회귀분석 모델의 계수는 최소자승법에 의해서 구해진다. 최소자승법은 추정오차를 이용하여 회귀분석 모델의 계수를 구하는 방법이며, 모델의 계수를 구하는 방법으로 범용적으로 사용되는 알고리즘이다. 최소자승법은 추정오차의 합이 최소가 될 때의 회귀분석 모델의 계수  $\alpha, \beta, \gamma$ 를 찾아

내는 기법이며, 수식은 다음과 같다[9].

$$\text{minimize } \sum_{i=1}^n (y_i - \alpha - \beta x_i - \gamma x_i^2)^2 \quad (5)$$

본 논문에서 다항회귀분석을 위한 입력 데이터( $x_i, y_i$ )는 다음과 같은 방식으로 만들어진다. 입력 데이터  $x_i$ 는 2.2절에서 소개한 온도효과와 전력 수요 간의 특징을 이용한 데이터 마이닝에 의해서 구해진 특수일 전 평일 4일의 값을 토대로 만들어진다. 데이터 마이닝을 통해 얻어진 평일 4일의 데이터를 각각  $m_1, m_2, m_3, m_4$ 라 하면 이 중 가장 큰 값(M)으로 정규화하여 이들의 평균으로 정한다[7].

$$x_i = \frac{m_1 + m_2 + m_3 + m_4}{4M} \quad (6)$$

$y_i$ 는 특수일 전력 수요량을  $m_5$ 라 하면 다음과 같이 표현된다[7].

$$y_i = \frac{m_5}{M} \quad (7)$$

2.3 특수일 전력 수요 예측 사례연구

본 절에서는 온도 효과를 고려한 데이터 마이닝을 적용한 다항회귀분석 알고리즘을 적용하여 특수일 전력 수요 예측 사례연구를 수행하였다. 기존 선형 퍼지 회귀분석과 제안된 방법 간 잔차와 전력 수요 예측 오차를 비교하였다. 기존 특수일 예측 방법으로 퍼지 선형회귀분석을 선택한 이유는 가장 최근에 연구된 내용이기 때문이며, 또한 퍼지 선형회귀분석법은 여타 다른 예측 방법들 보다 예측의 정확성이 우수하다는 것을 참고문헌[4]에서 확인할 수 있기 때문이다. 따라서 본 논문에서는 최근 연구된 전력 수요 예측 방법 중, 가장 높은 정확성을 보이는 퍼지 선형회귀분석법과 제안된 방법을 비교함으로써 본 논문의 알고리즘의 우수성을 입증하고자 한다.

회귀분석에서 잔차는 예측에 사용되는 회귀식이 기존 데이터의 특성을 얼마나 잘 표현하는 지를 보여주는 지표이다. 과거 데이터를 바탕으로 만들어지는 회귀식이 과거 데이터를 얼마나 잘 대표할 수 있는지 여부를 과거 데이터의 잔차를 통해 확인할 수 있다. 따라서 잔차가 작다는 것은 회귀식이 과거 데이터에 피팅(Fitting)이 잘 됐음을 의미하며, 다른 말로 표현하며 과거 데이터 특성을 비교적 정확히 반영하여 미래 데이터 예측이 가능하다는 것을 의미한다. 그림 4에서 보이는 잔차는 2006년 특수일 예측 시 사용된 과거 데이터들의 피팅 정도를 기존 퍼지 선형회귀분석법과 본 논문에서 제안된 방법으로 비교하여 나타난 것이다.

잔차 비교 시 제안된 방법인 다항회귀식을 사용하였을 경우 퍼지 선형회귀분석 시 때 보다 잔차가 작음을 그림 4에서 확인할 수 있다. 퍼지 선형회귀분석의 경우 잔차가 21% 넘는 경우가 생기는 반면, 제안된 방법에서는 모든 경우 잔차가 1% 이내에 존재한다. 따라서 이와 같은 잔차 결과는 본 논문의 제안된 방법이 기존 퍼지 선형회귀분석 보다 특

수일 전력 수요 예측에 적합하다는 것을 의미한다.

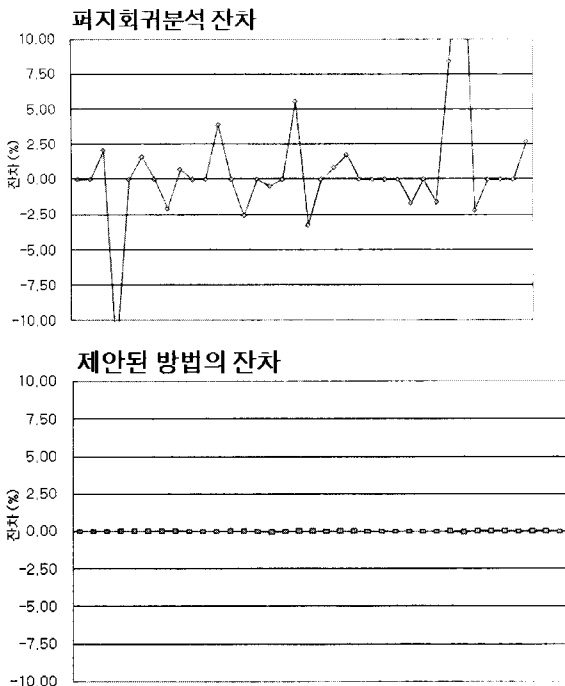


그림 4 전력 수요 예측 잔차 비교  
Fig. 4 Comparisons of Residuals

다음으로는 전력 수요 예측 오차를 비교하였다. 전력 수요 예측의 오차는 다음의 식(8)으로 구한다.

$$Error(\%) = \frac{|P^{Forecast} - P^{Actual}|}{P^{Actual}} \times 100 \quad (8)$$

여기서  $P^{Forecast}$ 는 예측 수요량이며,  $P^{Actual}$ 는 실제 수요량이다.<sup>1)</sup>

표 1에서는 본 논문에서 제안된 전력 수요 예측 방법과 퍼지 선형 회귀분석법의 2006년 특수일 전력 수요 예측 결과를 비교하였다. 제안된 방법을 적용 시, 특수일 전력 수요 예측에서 기존 퍼지 선형회귀 방법 대비 최대 5.71% 오차를 줄였으며, 전체 평균 오차를 기존 퍼지 선형회귀 방법의 5.03%에서 1.93%으로 낮추는 우수한 예측 정확성을 보였다. 제안된 방법을 사용하여 특수일 전력 수요를 예측한 표 1의 결과에서 개천절과 성탄절의 예측오차율이 다른 특수일 예측 오차율에 비해 비교적 크다. 이러한 결과를 보이는 이유는 2006년 개천절의 경우 추석 연휴기간에 포함되었기 때문이며, 성탄절의 경우 요일 타일이 월요일이어서 과거 데이터 중, 월요일 성탄절 데이터가 희귀한 경우이기 때문이다.

1) 실제 수요량은 한국전력거래소 홈페이지에 있는 데이터를 참조하였다.

표 1 2006년 특수일 전력 수요 예측 오차율 비교  
Table 1 Comparisons of Load Forecasting Performance for 2006 year

특수일	퍼지 선형 회귀(%)	제안된 방법(%)
신정	3.0860	1.2279
설날	2.8267	0.9313
삼일절	5.9572	0.8613
노동절	3.0289	1.0261
어린이날	3.8767	0.8630
석가탄신일	5.6260	1.8337
현충일	4.6838	0.4535
제헌절	3.9729	0.4202
광복절	4.6320	1.6606
추석	8.3935	2.6804
개천절	7.9990	5.4319
성탄절	6.2388	5.8286
평균	5.0268	1.9349

### 3. 결 론

특수일은 평일과는 다른 수요패턴으로 인해 수요 예측의 정확성이 크게 떨어진다. 이러한 현상을 개선하기 위해 본 논문에서는 결정계수를 이용한 데이터 마이닝을 적용한 과거 데이터 추출과 추출된 데이터를 이용한 다항회귀 예측 모델을 제안하였다. 다항회귀모델의 계수를 찾기 위해 최소자승법을 사용하였으며, 2006년 특수일을 대상으로 사례연구를 진행하였다. 사례연구 진행 결과 기존 방식인 퍼지 선형 회귀분석에 비해 특수일 전력 수요 예측 정확성이 향상됨을 입증하였다. 이러한 사례연구 결과는 전력 수요 예측을 위해 과거 데이터 추출 시 데이터 간의 상관관계, 즉 과거 전력 수요 패턴을 잘 보여주는 양질의 데이터 선택이 중요한 것을 말하며, 양질의 데이터의 기준으로 데이터 간의 상관관계를 표현하는 결정계수를 기준으로 나타낼 수 있음을 증명한다. 향후에는 퍼지 다항회귀분석법을 도입하여 특수일 전력 수요 예측의 정확도를 개선할 것이며, 특수일 뿐만 아니라 평일 전력 수요의 예측 혹은, 24시간 수요예측 등으로 개선된 알고리즘을 적용할 예정이다.

### 감사의 글

본 연구는 지식경제부의 지원으로 수행한 에너지자원인력양성사업의 연구결과입니다.

이 논문은 2008년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2008-331-D00195).

### 참 고 문 헌

[1] A.D. Papalexopoulos and T.C. Hesterberg, "A Regression-Based Approach to Short-Term System Load Forecasting," IEEE Trans. on Power Systems, vol. 4, no. 4, pp. 1535-1547, 1990.

- [2] A.G. Bakirtzis, et al., "A Neural Network Short-Term Load Forecasting Model for the Greek Power System," IEEE Trans. on Power Systems, vol. 11, no. 2, pp. 858-863, May 1996.
- [3] R. Lamedica, et al., "A Neural Network Based Technique for Short-Term Forecasting of Anomalous Load Periods," IEEE Trans. on Power Systems, vol. 11, pp. 1749-1756, Nov. 1996.
- [4] K.H. Kim, H.S. Youn, and Y.C. Kang, "Short-Term Load Forecasting for Special Days in Anomalous Load Conditions Using Neural Networks and Fuzzy Inference Method," IEEE Trans. on Power Systems ,vol. 15, no. 2, pp. 559-565, 2000.
- [5] H. Mori and A. Yuihara, "Deterministic Annealing Clustering for ANN-Based Short-Term Load Forecasting," IEEE Trans. on Power Systems ,vol. 16, no. 3, pp. 545-551, 2001.
- [6] K.H. Kim, H.S. Youn, and Y.C. Kang, "Short-Term Load Forecasting for Special Days in Anomalous Load Conditions Using Neural Networks and Fuzzy Inference Method," IEEE Trans. on Power Systems, vol. 15, no. 2, pp.559-565, 2000.
- [7] K.B. Song, Y.S. Baek, D.H. Hong, and G.S. Jang, "Short-Term Load Forecasting for the Holidays Using Fuzzy Linear Regression Method," IEEE Trans. on Power Systems ,vol.20, no.1, pp.96-101, 2005.
- [8] 위영민, 송경빈, 주성관, "결정계수 기반의 데이터 마이닝을 이용한 특수일 최대 전력수요 예측," 대한전기학회 하계학술대회, 2008
- [9] 류근관, "통계학", 법문사, 2003

## 저 자 소 개



### 위 영 민 (魏 永 民)

1980년 4월 7일생. 2005년 중앙대학교 공과대학 전자전기공학부 졸업(공학사). 2005년 삼성전자 엔지니어 근무. 현재, 고려대학교 공과대학 전자전기공학과 석사과정.



### 송 경 빈 (宋 敬 彬)

1963년 9월 15일생. 1986년 연세대학교 공과대학 전기공학과 졸업(공학사). 1988년 동 대학원 전기공학과 졸업(공학석사). 1995년 텍사스 A&M University 전기공학과 졸업(박사). 현재, 숭실대학교 전기공학부 부교수.



### 주 성 관 (朱 成 官)

1971년 7월 6일생. 1995년 고려대학교 공과대학 전기공학과 졸업(공학사). 1999년 University of Washington 졸업(공학석사). 2004년 동 대학원 전기공학과 졸업(공학박사). 현재, 고려대학교 전기전자전파공학부 조교수.