2008년도 대한전기학회 하계학술대회 논문집 2008. 7. 16 - 18

결정계수 기반의 데이터 마이닝을 이용한 특수일 최대 전력 수요 예측

위영민*, 송경빈**, 주성관*
고려대학교*, 숭실대학교**

Load Forecasting for the Holidays Using a Data mining with the Coefficient of Determination

Young-Min Wi*, Kyung-Bin Song**, Sung-Kwan Joo*
Korea University*, Soongsil University**

Abstract - 본 논문에서는 특수일 전력 수요 예측을 위한 알고리즘을 제시하였다. 논문에서 제안하는 전력 수요 예측 알고리즘은 데이터 마이닝을 이용한 데이터 전처리 부분과 전처리된 데이터를 사용하여 특수일 수요를 예측하는 다항 회귀분석 부분으로 나누어 진다. 데이터 전처리에서는 전력 수요 예측을 위한 과거 데이터 중에 과거 특수일 수요의 패턴을 잘 보여주는 데이터를 찾기 위해 온도와 수요의 관계를 이용한다. 데이터 마이닝의 기준으로 결정계수를 사용하였으며, 알고리즘은 k-nearest neighbor 절차를 사용하였다. 또한 제안된 기법은 2006년 특수일 전력 수요 예측을 통하여 기존 논문의 결과와 비교 분석하여 기존 방식 대비 특수일 전력 수요예측 관련 우수성을 검증하였다.

1. 서 론

전력 수요 예측은 전력 시스템(안정도, 유지, 보수, 투자) 뿐만 아니라 전력 시장에서 중요한 지표로 사용된다. 또한 최근 진행 중인 다양한 연구에서 기본이면서 가장 중요한 요소로 작용된다. 예를 들어 장기 수급 계획, 분산전원, 수요 반응 등 매우 다양한 분야에서 기초 자료로 사용된다. 하지만 전력 수요는 그 특성이 비선형적이기 때문에 예측하는 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. 이러한 모든 요소를 감안하여 전력 수요를 예측하기 위한 시스템을 계획하기는 현실적으로 어려운 일이다.

전력 수요 예측 기법은 매우 다양하게 발전되어 왔다. 초창기 시계열법[1], 회귀분석법 등이 연구 되었으며, 국·내외로 전문가 지식 기반의 인공지능기법[2]-[4]과 퍼지 방법[5] 통한 예측 방법이 연구되어 지고 있다. 또한 최근에는 기존 방법들을 결합한 하이브리드 방식 알고리즘에 대한 연구가 활발히 진행 중이다. 현재까지 진행된 연구 내용에서 알 수 있듯 평일 대상의 전력 수요 예측은 대부분 높은 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.

정확한 전력 수요 예측을 위해서는 비선형적인 요소를 얼마나 정확히 제거하여, 전력 수요 특성을 잘 나타내는 양질의 데이터를 사용하느냐가 중요한 점으로 작용된다. 본 논문에서 양질의 데이터 즉, 전력 수요의 패턴을 잘 표현하는 데이터를 얻기 위해 통계학에서 사용하는 결정계수를 이용한 데이터 마이닝을 사용하였다. 또한 결정계수를 이용하여 걸러진 데이터를 다항회귀분석의 입력변수로 사용하여 기존 방식 보다 정확한 특수일 전력 수요 예측 결과를 얻을 수 있었다.

본 논문에서는 특수일 수요 예측의 정확성을 평일 수준으로 끌어올리긴 위한 알고리즘을 제시하고 제시된 방법에 따른 시험 결과를 예시하였다.

2. 본 론

2.1 전력 수요 예측

한국 전력 거래소에 제공하는 예측 오차를 확인하면 특수일과 평일(평일, 주말)의 편차가 큼을 쉽게 확인할 수 있다. 이러한 현상의 가장 큰 이유는 평일 대비 특수일이 가진 과거 데이터의 희소성이 근본적 원인이다. 정확한 예측을 하기 위해서는 많은 데이터가 있어야 한다. 하지만 데이터가 많아도 양질의 데이터가 부족하다면 절대적인 과거 데이터의 양은 무의미하다. 따라서 우리는 과거의 데이터 중에 과거 전력 수요 패턴을 잘 보여주는 양질의 데이터를 얻기 위한 알고리즘이 필요하다. 본 논문은 이를 위해 결정계수를 이용하여 과거의 수요 데이터의 온도 영향성을 분석하여 시뮬레이션을 위한 과거 데이터 집합을 구성하였다. 또한 이렇게 구성한 과거 데이터 집합을 이용하여 다항회귀분석을 통해 특수일 전력 수요 예측을 진행하였다.

2.2절에서는 온도효과를 이용하여 과거 데이터 선택을 위한 결정계수를 사용한 데이터 마이닝 알고리즘을 제시한다. 2.3절에서는 선택된 데이터를 사용한 다항회귀분석법을 소개한다. 2.4절에서는 제안된 알고리즘을 이용하여 2006년 특수일 전력 수요 예측을 통하여 제안된 알고리즘을 비교 검증하였다.

2.2 결정계수를 사용한 특수일 전력 수요 데이터 마이닝

정확한 전력 수요 예측 모델을 위해서는 과거 수요 패턴을 잘 보여주는 과거 데이터가 필요하다. 본 논문에서 제안하는 알고리즘은 특수일과 평일 수요 데이터의 관계를 이용하여 미래 특수일 수요를 예측한다. 정확한 특수일 예측을 위해 과거 특수일이 가진 수요의 특성을 잘 나타내는 과거의 데이터 선택이 필수적이다. 제안된 알고리즘은 온도를 이용하여 특수일과 평일 수요 데이터의 관계의 유사도를 판단하여 평일 과거 데이터를 찾아낸다. 유사도를 측정하기 위해 통계학에서 사용되는 결정계수를 이용한다. 결정계수는 두 변수 간의 설명력이 최대일 때 1이 되고, 최소일 때 0이 된다. 결정계수는 0과 1사이에서 존재한다. 결정계수는 수식 (1)-(3)에 의해서 구해진다. 독립변수 x와 종속변수 y 로 이루어진 n개의 회귀분석 샘플이 존재할 때, 결정계수는

$$\text{결정계수}(R^2) = 1 - \frac{SSE}{SST} \tag{1}$$

$$\text{총제곱합}(SST) = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \tag{2}$$

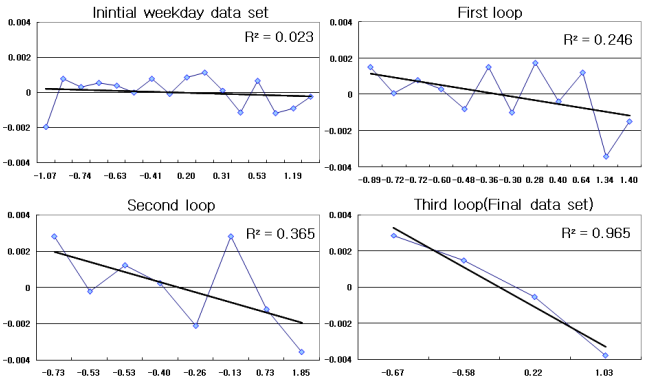
$$\text{오차제곱합}(SSE) = \sum_{i=1}^{n} \varepsilon_i^2. \tag{3}$$

여기서 Y_i는 i번째 종속 변수, $\overline{Y}$는 Y_i의 평균이며. ε_i는 회귀분석 오차이다.

본 논문에서는 특수일 전력 수요 예측을 위해 다항회귀분석을 사용한다. 회귀분석의 독립변수 x로는 특수일 이 전의 평일 전력 수요 데이터의 평균을 사용하며, 종속변수로는 특수일 전력 수요 데이터를 사용한다. 과거 평일 데이터 중 x를 선택하기 위해 본 논문에서는 특수일 최고 기온과 같은 온도를 갖는 평일 데이터를 추출한다. 제안된 알고리즘은 평일 전력 수요 데이터 선택을 위해 결정계수를 사용하며, 아래에서 소개되는 절차에 의해 추출된 평일 데이터 중에 특수일 온도의 전력수요 패턴을 보여줄 수 있는 데이터를 선택하게 된다.

a) 평일 데이터를 온도와 전력수요에 관해서 정규화 한다.
b) 정규화된 평일 데이터의 결정계수를 구한다.
c) 결정계수가 임계값보다 작을 경우, b)에서 구한 결정계수가 작은 데이터 4개를 데이터 집합에서 없앤다. 만약 결정계수가 임계값보다 큰 경우 f)로 점프한다.
d) c)에서 남은 데이터를 가지고 다시 결정계수를 계산한다.
e) c)와 d)를 데이터가 4개 남을 때까지 반복한다.
f) 남은 데이터 중 잔차가 작은 상위 4개의 데이터를 선택한다.

그림 1은 결정계수를 이용한 데이터 마이닝을 보여주는 그림이다. 그래프에서 x, y축은 각각 정규화된 온도와 전력 수용이다.



<그림 1> 결정계수를 이용한 데이터 마이닝

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 단순회귀분석이 보다 다항회귀분석을 주로 이용한다. 일반적인 2차 다항회귀 모델은 다음의 수식(4)과 같다.

$$y = \alpha + \beta x + \gamma x^2 \tag{4}$$

여기서 α, β, γ는 2차 회귀분석 모델 계수이다.

2차 회귀분석 모델의 계수는 최소자승법에 의해서 구해진다. 최소자승법은 추정오차를 이용하여 회귀분석 모델의 계수를 구하게 된다. 즉, 최소자승법은 추정오차의 합이 최소가 될 때의 회귀분석 모델의 계수 α, β, γ를 찾아내는 기법이며, 수식은 다음과 같다[6].

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

본 논문에서 다항회귀분석을 위한 입력 데이터(X_i, Y_i)는 다음과 같은 방식으로 만들어진다. 입력 데이터 X_i는 2.2절에서 소개한 온도효과와 전력 수요간의 특징을 이용한 데이터 마이닝에 의해서 구해진 특수일 전 평일 4일의 값을 토대로 만들어 진다. 데이터 마이닝을 통해 얻어진 평일 4일의 데이터를 각각 m_1, m_2, m_3, m_4라 하면 이 중 가장 큰 값(M)으로 정규화하여 이들의 평균으로 정한다[5].

$$x_i = \frac{m_1 + m_2 + m_3 + m_4}{4M} \tag{6}$$

Y_i는 특수일 수요량을 m_1라 하면 다음과 같이 표현된다[5].

$$y_i = \frac{m_5}{M} \tag{7}$$

2.4 특수일 전력 수요 예측 사례연구

본 절에서는 온도 효과를 고려한 데이터 마이닝을 적용한 다항회귀분석 알고리즘을 적용하여 사례연구를 수행하였다. 수요예측의 오차는 다음의 식(5)으로 구한다.

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

여기서 $P^{Forecast}$는 예측 수요량이며, P^{Actual}는 실제 수요량이다.[1)]

표 1에서는 2006년 특수일 전력 수요 예측을 제안된 방법으로 테스트를 진행하였다.

<표 1> 2006년 특수일 전력 수요 예측 오차 비교

특수일	퍼지 선형 회귀	제안된 방법
양력설날	3.0860	1.2279
음력설날	2.8267	0.9313
삼일절	5.9572	0.8613
노동절	3.0289	1.0261
어린이날	3.8767	0.8630
부처님	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

제안된 방법을 적용 시, 특수일 전력 수요 예측에서 기존 논문 대비 최대 5.71% 오차를 줄였으며, 전체 평균 오차를 1.93% 낮추는 우수한 예측 정확성을 입증하였다.

3. 결 론

특수일은 평일과는 다른 수요패턴으로 인해 수요 예측의 정확성이 크게 떨어진다. 이러한 현상을 개선하기 위해 본 논문에서는 온도 효과를 고려한 데이터 마이닝을 적용한 다항회귀분석 모델을 제안하였다. 다항회귀모델의 계수를 찾기 위해 최소자승법을 사용하였으며, 2006년 특수일을 대상으로 사례연구를 진행하였다. 사례연구 진행 결과 기존 방식인 퍼지 선형회귀분석에 비해 특수일 전력 수요 예측 정확성이 향상됨을 입증하였다.

감사의 글

본 연구는 지식경제부의 전력산업 고급인력양성사업을 통한 지원으로 수행되었음.

[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] 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.
[6] 류근관, “통계학”, 법문사, 2003

1) 실제 수요량은 전력거래소 홈페이지 데이터를 사용하였다.