

다중회귀모형을 이용한 월간 전력수요 예측기법

문지훈, 김용성, 박진웅, 황인준
고려대학교 전기전자공학과

e-mail:{johnny89, kys1001, timeless, ehwang04}@korea.ac.kr

Monthly Electric Load Forecasting Method Using Multiple Regression Model

Jihoon Moon, Yongsung Kim, Jinwoong Park and Eenjun Hwang
School of Electrical Engineering, Korea University

요 약

전력수요 예측은 설비투자, 수급 안정, 구매전력비 등에 직결되는 중요한 요소이며 국가 경제에 미치는 영향이 크다. 특히 인구가 밀집한 대도시의 경우 정치, 교육, 문화, 경제적 활동들이 전력사용과 밀접한 연관이 있어 안정적인 전력공급을 위한 정확한 전력수요 예측이 필요하다. 최근 평균기온 및 국내총생산을 독립변수로 활용하여 다중회귀모형을 구성한 연구가 전국 단위 전력수요 예측에 유용한 결과를 보여주었다. 하지만 좀 더 작은 단위 지역의 전력수요를 예측할 때에는 지역마다 제반 여건에 따른 전력사용 용도가 다르므로, 그 지역의 전력수요와 상관관계가 높은 다른 변수들을 함께 고려해야 할 필요가 있다. 본 논문은 서울시 자치구별 월 단위 전력수요 예측을 위하여 과거 전력수요량을 독립변수, 평균기온, 지역내총생산, 자치구별 인구, 세대수, 지하철 승·하차 인원을 종속변수로 설정한 다중회귀모형을 구성하였다. 이를 기반으로 다양한 실험을 통해 자치구별 월간 전력수요 예측을 진행하였으며, 그 결과 이전보다 향상된 정확도를 얻을 수 있었다.

1. 서론

전력계통의 최적 운영계획을 위해서는 사전에 전력수요를 정확하게 예측해야 할 필요가 있다. 전력수요의 종류는 사용하는 용도에 따라 크게 가정용, 상업용, 공공용으로 나누어진다. 지역의 전력수요는 지역마다 제반 여건이 다르므로 전력을 사용하는 용도가 다르다. 지역별로 정확한 전력수요를 예측하려면 전력수요와 상관관계가 높은 다른 요인들을 반영하여 제반 여건을 고려해야 한다.

대도시의 경우 시장, 상업, 전문 상가를 중심으로 발달한 상업업을 비롯해 운수, 통신, 금융 등의 산업이 집중되어 있어, 잘못된 전력수요 예측으로 인한 정전 발생 시에 경제적 손실은 견잡을 수 없이 클 것이다. 그러므로 대도시에 대한 전력수요 예측은 특히 중요하다 할 수 있다. 대표적 대도시 지역인 서울의 경우, 자치구의 제반 여건에 따라 서비스업(상업용)이 활성화된 지역과 주거인구(가정용)가 많은 지역이 전력수요와 높은 상관관계를 가진다[1].

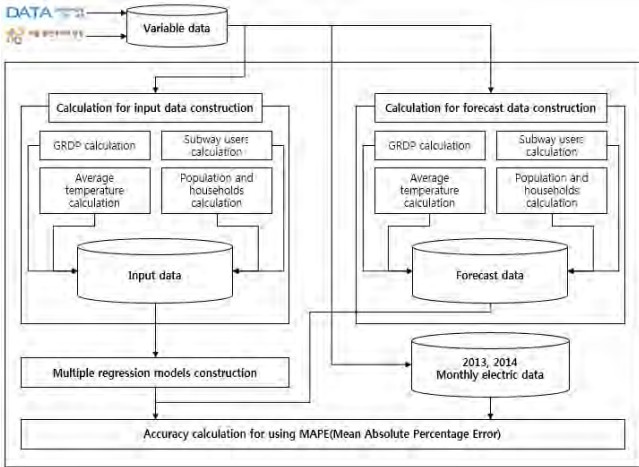
본 논문은 다중회귀모형을 이용한 서울특별시 자치구별 월간 전력수요 예측기법을 제안한다. 다중회귀모형에서 구성된 종속변수는 자치구별 월간 전력수요량이고, 독립변수는 평균기온, GRDP(Gross Regional Domestic Product, 지역내총생산), 자치구별 인구, 세대수, 지하철 승·하차 인원이다. 자치구별 인구와 세대수를 가지고 가정용 전력수요를 고려하였다. 유동인구가 많은 지역과 그 지역의 서비스업(상업용) 활성화는 높은 상관관계를 가지므로[6], 자치

구별 지하철 승·하차 인원을 통해 상업용 전력수요를 고려하였다. 제안한 다중회귀모형은 MAPE(Mean Absolute Percentage Error, 평균 절대 백분율 오차)를 적용하여, 선행 연구[4]에서 제안된 평균기온, GRDP를 독립변수로 구성한 다중회귀모형과 전력수요 예측 결과를 비교하였다.

2. 관련 연구

전력수요의 예측 정확성을 개선하기 위한 전력수요 예측 기법에 관한 연구가 지속해서 이루어져 왔다. [2]는 특수일의 특성에 맞는 입력인자를 선택하고, ELM(Extreme Learning Machine)을 사용하여 설날과 추석 연휴에 대한 최대 전력수요 예측을 수행하였다. [3]은 PCA(Principal Component Analysis) 기법을 적용하여 예측 모델의 입력변수로 활용되는 변수들을 선택하고, 특징 추출하여 SVR(Support Vector Regression)을 적용한 월간 전력수요 예측기법을 제안하였다. [4]는 전력수요와 관계가 큰 기온과 GDP(Gross Domestic Product)를 주 특성에 맞게 변환하여 다중회귀모형을 구축한 뒤, 주 최대 전력수요를 예측하였다. [5]는 하계 특수경부하기간의 추세를 세 가지(증가, 수평, 감소) 패턴으로 분류하여 각 추세에 대한 전력수요 예측 알고리즘을 제시하였다. 이러한 전력수요 예측기법들은 우수한 성능을 보인다. 하지만, 작은 단위 지역의 전력수요를 예측하기 위해서는 지역의 전력사용 용도를 고려한 모형의 개발 및 개선의 필요성이 있다.

3. 다중회귀분석을 이용한 전력수요 예측 시스템



(그림 1) 전력수요 예측 시스템 구성도

본 논문에서 제안하는 전력사용 예측 시스템 구성은 그림 1과 같다. 먼저 ‘서울 열린 데이터 광장’과 ‘공공데이터 포털’을 통해 다중회귀모형 구성에 필요한 데이터를 수집한다. 이를 바탕으로 다중회귀모형에 필요한 입력 데이터 값을 구성한다. 다음으로 구성된 다중회귀모형에 입력할 예측 데이터값을 생성하고, 그 예측값을 계산한다.

3.1 입력 데이터 구성

서울시 자치구별 월간 전력수요를 예측하기 위해 구성되는 다중회귀모형에서 종속변수는 자치구별 월간 전력수요 량이고, 독립변수는 평균기온, GRDP, 자치구별 인구, 세대수, 지하철 승·하차 인원이다. 본 논문의 목적은 월간 전력수요를 예측하는 것이어서, 독립변수들을 월 단위로 변환해야 할 필요가 있다.

1) 월 실측 기온 계산

기상청에서 제공하는 기온 데이터는 평균기온, 최고기온, 최저기온이 있다. 본 논문은 기존 연구[2-4]에서 전력수요 예측 변수로 사용된 평균기온을 사용한다. 기온 데이터에서 주어지는 일, 월간 평균기온 중 월간 평균기온을 사용하였다. 우리나라 전력수요는 여름과 겨울에 높게 나타난다. 그러나 겨울 기온은 낮고 여름 기온은 높아 기온 값과 전력수요와 상관관계가 크지 않다. 기온 값과 전력수요의 상관관계를 높이는 방안으로 식 (1)을 적용하였다.

$$ATC_{i,j} = |AT_{totalaverage} - AT_{i,j}| \quad (1)$$

AT 는 Average Temperature의 약자로 서울시 월간 평균기온을 의미한다. ATC (Average Temperature Conversion)는 서울시 월간 평균기온과 전력수요의 상관관계를 높이기 위해 제안한 보정된 값이다. $AT_{totalaverage}$ 는 다중회귀모형을 구성한 시점인 2005년부터 2012년까지의 서울시 전

체 월간 평균기온의 평균값이다. $AT_{i,j}$ 는 $ATC_{i,j}$ 를 구하고자 하는 j 년 i 월의 평균기온을 말한다.

<표 1> 평균기온(AT), 보정된 값(ATC)과 서울시 전력수요의 피어슨 상관계수

	평균기온	보정된 값
2005	0.174	0.734
2006	0.138	0.779
2007	0.343	0.786
2008	0.020	0.860
2009	0.042	0.797
2010	0.103	0.815
2011	-0.184	0.852
2012	-0.085	0.771

표 1은 연도별 월 평균기온, 제안한 보정된 값과 서울시 전력수요의 상관관계를 피어슨 상관계수로 나타낸 것이다. 일반적으로 1과 가까울수록 양의 상관관계가 있으며 계수의 값이 0.7 이상이면 강한 양적 선형관계를 나타낸다[9]. 본 논문에서 제안한 보정된 값을 적용하면, 계절의 특성을 반영하여 더욱 효과적인 다중회귀모형을 구성할 수 있다.

2) 월 실측 GRDP 계산

GRDP는 Gross Regional Domestic Product의 약어이며 지역내총생산을 의미한다. 통계청 공공데이터에서 제공하는 연간 GRDP를 가지고 월간 GRDP를 추정하는 과정이 필요하다. 월간 GRDP를 구하는 방식은 공공데이터에서 월간 제공되는 생산지수를 가지고 식 (2)를 적용하였다.

$$GRDP_{i,j} = \frac{Production\ Index_{i,j}}{Production\ Index_j} \times GRDP_j \quad (2)$$

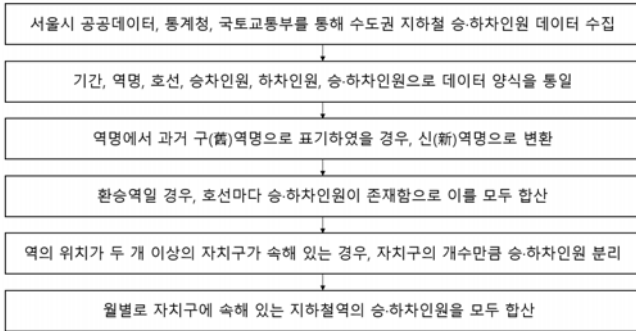
$Production\ Index_j$ 는 월간 생산지수가 속해 있는 j 년도의 전체 월간 생산지수 합을 나타낸다. $Production\ Index_{i,j}$ 는 $GRDP_{i,j}$ 를 계산하는 j 년 i 월에 해당하는 생산지수이다. $GRDP_j$ 는 $GRDP_{i,j}$ 가 속한 j 년도의 GRDP다.

3) 월 실측 자치구별 인구, 세대수 계산

상주인구가 많은 지역과 그 지역의 가정용 전력수요는 높은 상관관계를 가진다[1]. 본 논문은 이러한 제반 여건을 고려하기 위해 자치구별 월간 인구를 계산하였다. 서울시 공공데이터에서 제공하는 인구가동 데이터를 가지고, 2004년 12월에 측정된 자치구 인구 기준으로 2005년부터 2012년까지 월간 자치구 인구를 구하였다. 세대수의 경우 월간 데이터를 제공하지 않고, 연도별 데이터를 제공한다. 본 논문은 세대수의 경우 연도별 데이터를 가지고 다중회귀모형을 구성하였다.

4) 월 실측 지하철 승·하차 인원 계산

유동인구가 많은 지역과 그 지역 서비스업의 활성화는 높은 상관관계를 가진다[6]. 본 논문은 지하철 승·하차 인원을 통해 상업용 전력수요를 고려하였다. 서울시는 환승역 중복을 제외하여 총 291개의 역(2016년 2월을 기준)이 있다. 위키백과[8] 기준으로 자치구별로 속한 지하철역을 정하였다. 그림 2에 제시한 과정을 통해 자치구별 지하철 승·하차 인원을 구하였다.



(그림 2) 자치구별 지하철 승·하차 인원 계산 과정

3.2 예측 데이터 구성

2005년부터 2012년까지의 공공데이터를 가지고, 구성된 다중회귀모형에 입력할 2013년, 2014년 예측 데이터값을 다음과 같은 기준으로 구하였다.

1) 월 예측 기온 계산

월 예측 기온은 2012년 기상청에서 출간한 ‘한반도 기후 변화 전망보고서’[10]를 기준으로 예측하였다. [10]에서 예측한 연간 기온 상승률을 가지고, 2012년 월간 평균기온에서 2013년도 월간 평균운도를 구하였다. 앞서 예측한 2013년 월간 평균기온에서, [10]가 예측한 기온 상승률을 반영하여 2014년도 월간 평균기온을 구하였다.

2) 월 예측 GRDP 계산

매년 7월에 한국은행은 내년도 경제를 전망하는 ‘한국은행 경제전망’[11]을 출간한다. 2012년 7월에 출간한 [11]이 전망한 GDP(Gross Domestic Product, 국내총생산) 증감률을 통해, 2012년 월간 GRDP를 가지고 2013년 월간 GRDP를 예측하였다. 2013년 7월에 출간한 [11]이 전망한 GDP 증감률을 통해, 앞서 예측한 2013년 월간 GRDP를 가지고 2014년 월간 GRDP를 예측하였다.

3) 월 예측 인구, 세대수 계산

월 예측 인구, 세대수는 서울특별시에서 2012년 11월 26일에 작성한 ‘서울 「2012~2034년 서울 인구 및 가구구조 변화전망」 통계’[12]를 가지고 예측하였다. 2013년 자치구 세대수와 월간 인구는 [12]에서 예측한 2013년도 인구와 세대수의 증감률을 가지고, 2012년 자치구 세대수와 월간 인구를 바탕으로 예측하였다. 2014년 자치구 세대수와 월

간 인구는 [12]에서 예측한 2014년도 인구와 세대수의 증감률을 통해, 앞서 예측한 2013년도 예측값을 가지고 예측하였다.

4) 월 예측 지하철 승·하차 인원 계산

월 예측 지하철 승·하차 인원인 경우, 2009년 8월부터 2012년 12월까지 자치구 지하철 승·하차 인원을 가지고 단순선형회귀모형을 구성하였다. 예측값의 정확성을 위해 2009년 7월 24일 9호선이 개통되는 시점을 기준으로, 9호선 승·하차 인원이 반영된 2009년 8월을 선정하였다. 이를 통해 2013년 1월부터 2014년 12월까지의 자치구별 월간 지하철 승·하차 인원을 예측하였다.

3.3 다중회귀분석을 이용한 예측 모형

회귀분석은 관찰된 연속형 변수들에 대해 독립변수와 종속변수 사이의 상관관계를 나타내는 선형 관계식을 구하는 기법이다. 다중회귀모형에서 종속변수 Y 를 설명하는데 독립변수 x_1, \dots, x_k 를 이용하여 수식을 구성한다.

$$Y_1 = \beta_0 + \beta_1ATC + \beta_2GRDP + \beta_3Population + \beta_4SubwayUsers + \epsilon_1 \quad (3)$$

식 (3)의 경우, 독립변수를 월간으로 평균기온, GRDP, 자치구 인구, 자치구 지하철 승·하차 인원을 가지고 구성된 다중회귀모형이다. Y_1 은 자치구 월간 전력수요, ATC 는 월간 평균기온의 보정된 값, $GRDP$ 는 월간 GRDP, $Population$ 은 자치구 월간 인구, $SubwayUsers$ 는 자치구 월간 지하철 승·하차 인원이다. ϵ_1 은 오차항, β_0 는 상수항, $\beta_1, \beta_2, \beta_3, \beta_4$ 은 각 변수의 회귀계수를 의미한다.

$$Y_2 = \beta_0 + \beta_1ATC + \beta_2GRDP + \beta_3Households + \beta_4SubwayUsers + \epsilon_2 \quad (4)$$

식 (4)의 경우, 독립변수를 월간으로 평균기온, GRDP, 자치구 세대수, 자치구 지하철 승·하차 인원을 가지고 구성된 다중회귀모형이다. $Y_2, ATC, GRDP, SubwayUsers, \epsilon_2, \beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ 은 식 (3)과 같은 의미가 있다. $Households$ 는 연도별 자치구 세대수이다.

4. 실험 및 결과

본 연구는 서울 열린 데이터 광장과 공공데이터포털을 통해 다중회귀모형 구성에 필요한 데이터를 수집하였다. 실험은 R-3.2.3을 기반으로 RStudio에서 진행하였다.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{\text{실제 전력수요} - \text{예측 전력수요}}{\text{실제 전력수요}} \right| \quad (5)$$

전력수요 예측의 정확성은 전력수요 예측 오차율을 통해 알 수 있다. 전력수요 예측 오차는 실제 전력수요와 예측 전력수요의 차이를 의미하며, 식 (5)를 적용하여 나타났다.

<표 2> 제안한 다중선형회귀모형을 이용한 평균 절대 백분율 오차(MAPE)

기간	선행 연구	제안한 기법 1	제안한 기법 2	제안한 기법 3
2013	7.3%	4.7%	4.0%	3.6%
2014	8.9%	7.9%	7.4%	6.9%
전체	7.6%	6.8%	5.7%	5.3%

MAPE는 Mean Absolute Percentage Error의 약어이며 평균 절대 백분율 오차를 의미한다.

n 은 전력수요 예측 기간인 2013년 1월부터 2014년 12월 까지 총 24개월을 말한다. 서울특별시의 자치구는 총 25개로 최종적으로 구한 n 의 값은 600이다. 2013년과 2014년 그리고 전체 MAPE의 값을 표 2에 나타냈다.

선행 연구[4]는 평균기온과 GRDP를 독립변수로 구성한 다중회귀모형이다. 제안한 기법 1은 식 (3)을 가지고 구성된 다중회귀모형이며, 제안한 기법 2는 식 (3)을 월간으로 분류하여 구성된 다중회귀모형이다. 예를 들어 2013년, 2014년 1월 전력수요 예측을 수행하기 위해, 2005년부터 2012년까지 각 연도의 1월 데이터만을 가지고 구성된 다중회귀모형이다. 제안한 기법 3은 식 (4)를 가지고 월간으로 분류하여 구성된 다중회귀모형이다. 제안한 기법들은 선행 연구보다 MAPE의 값이 적음을 확인하였다. 특히 제안한 기법 3은 2013년도는 3.7%, 2014년도는 2%, 전체는 2.3% 오차율을 줄여, 선행 연구보다 30.3%가 향상된 효과적인 전력수요 예측기법임을 증명하였다. 이는 가정용 전력수요가 인구보다 세대수에 의한 영향이 더 크다는 것 [7]을 확인할 수 있었다.

예측 결과의 오차율이 다소 높은 이유는 2013년을 기점으로 서울시에서 실시한 ‘원전하나줄이기’ 사업[13]이 있다. 이 사업으로 기업, 가정, 학교 등의 적극 참여와 신재생에너지 생산 확대로 전력수요가 감소하는 추세를 보였다. 본 논문은 최신 경향을 반영하기 위해 2013년과 2014년도 자치구별 월간 전력수요 예측을 하였다. 그러나 전력수요가 상승하는 시점에서 실험을 진행하였다면 MAPE의 값이 더 적었을 것으로 판단된다.

5. 결론 및 향후 연구

본 연구는 전력수요 예측의 정확성 개선을 위해, 서울시의 전력사용 용도를 고려한 다중회귀모형을 구성하였다. 기존 연구에서 다중회귀모형으로 구성된 독립변수인 평균기온, GRDP와 서울시의 제반 여건을 고려한 인구, 세대수, 지하철 승·하차 인원을 추가하여 기존 연구보다 향상된 전력수요를 예측하였다. 특히 평균기온, GRDP, 세대수, 지하철 승·하차 인원으로 구성된 다중회귀모형은 선행 연구보다 7.6%에서 5.3%로, 2.3%의 오차율을 줄여 30.3%가 향상된 효과적인 전력수요 예측기법임을 입증하였다.

향후 전력수요와 상관관계가 높은 다른 변수들을 고려하여 연구를 진행할 것이다. 또한 다양한 지역을 가지고 본 논문에서 제안한 기법의 범용성 여부를 확인할 계획이다.

6. Acknowledgements

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술연구원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 20152010103060)

참고문헌

- [1] 서울연구원 서울경제분석센터, “서울시 구별 전력 에너지 소비량,” 서울경제, 2013.
- [2] P. S. Ji and J. Y. Lim, “Development of Peak Power Demand Forecasting Model for Special-Day using ELM,” The Transactions of the Korean Institute of Electrical Engineers, Vol. 64, No. 2, pp. 74-78, 2015.
- [3] H. R. Lee and H. J. Shin, “Electricity Demand Forecasting based on Support Vector Regression,” IE interfaces, Vol. 24, No. 4, pp. 351-361, 2011.
- [4] H. W. Jung, S. Y. Kim, and K. B. Song, “Weekly Maximum Electric Load Forecasting Method for 104 Weeks Using Multiple Regression Models,” The Transactions of the Korean Institute of Electrical Engineers, Vol. 63, No. 9, pp. 1186-1191, 2014.
- [5] J. D. Park and K. B. Song, “Short-Term Load Forecast for Summer Special Light-Load Period,” The Transactions of the Korean Institute of Electrical Engineers, Vol. 62, No. 4, pp. 482-488, 2013.
- [6] J. M. Yun and D. J. Choi, “Geographically Weighted Regression on the Characteristics of Land Use and Spatial Patterns of Floating Population in Seoul City,” Journal of the Korean Society for Geospatial Information Science, Vol. 23, No. 3, pp. 77-84, 2015.
- [7] S. Y. Kim, “The Effect of Changes of the Housing Type on Long-Term Load Forecasting,” The Transactions of the Korean Institute of Electrical Engineers, Vol. 64, No. 9, pp 1276-1280, 2015.
- [8] 위키백과, 수도권 전철 [Internet], https://ko.wikipedia.org/wiki/수도권_전철
- [9] 위키백과, 상관 분석 [Internet], <https://ko.wikipedia.org/wiki/상관분석>
- [10] 기상청, “한반도 기후변화 전망보고서,” 2012.
- [11] 한국은행, “한국은행 경제전망,” 2012-2013.
- [12] 서울통계, “서울 인구 및 가구구조 전망 통계,” 2012.
- [13] 서울정책아카이브, “원전하나줄이기,” 2015.