

데이터 군집화를 이용한 전력 사용량 예측 기법

박진웅, 문지훈, 김용성, 황인준

고려대학교 전기전자공학과

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

Electric Power Consumption Forecasting Method using Data Clustering

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

요 약

최근 에너지 효율을 최적화하는 차세대 지능형 전력망인 스마트 그리드 시스템(Smart Grid System)이 국내외에 널리 보급되고 있다. 그로 인해 그리드 시스템의 효율적인 운영을 위해 적용되는 EMS(Energy Management System) 기술의 중요성이 커지고 있다. EMS는 에너지 사용량 예측의 높은 정확성이 요구되며, 예측이 정확하게 수행될수록 에너지의 활용성이 높아진다. 본 논문은 전력 사용량 예측의 정확성 향상을 위한 새로운 기법을 제안한다. 구체적으로, 먼저 사용량에 영향을 미치는 환경적인 요인들을 분석한다. 분석된 요인들을 적용하여 유사한 환경을 가지는 전력 사용량 데이터의 사전 군집화를 수행한다. 그리고 예측 일에 관련된 환경 정보와 가장 유사한 군집의 전력 사용량 데이터를 기반으로 전력 사용량을 예측한다. 제안하는 기법의 성능을 평가하기 위해, 다양한 실험을 통하여 일간 전력 사용량을 예측하고 그 정확성을 측정하였다. 결과적으로, 기존의 기법들과 비교했을 때, 최대 52.88% 향상된 전력 사용량 예측 정확성을 보였다.

1. 서론

최근 에너지 효율을 최적화하는 스마트 그리드 시스템(Smart Grid System)에 대한 수요가 증대되면서 이에 대한 연구 및 개발의 필요성이 커지고 있다. 스마트 그리드의 핵심 요소 중 하나인 마이크로 그리드는 전기 에너지를 자급자족할 수 있으며, 에너지 저장 시스템(ESS: Energy Storage System)과 결합하여 생성된 에너지를 저장할 수 있는 형태를 가진다[1]. 이는 전력공급회사에 대한 의존도를 낮추고 자체 전력망 내에서 전기 수요를 충당할 수 있도록 전력을 생산한다. 발전 시스템의 스케줄링은 전력 사용량 예측치를 기반으로 수행된다. 그러므로 전력사용량 예측의 정확성은 효율적인 발전 시스템 운용에서 가장 중요한 부분이다.

최근 대한민국은 여름이 길어지고 겨울이 짧아지는 이상기온 현상이 빈번히 발생하고 있다. 선행 연구[4, 5]에서 적용된 변수인 계절 정보를 가지고 예측을 수행하는 경우에는 정확성이 다소 떨어지는 문제점이 있다.

본 연구는 선행 연구보다 스마트 그리드 시스템에 적합한 전력 사용량 예측의 새로운 기법을 제안한다. 먼저 전력 사용량에 영향을 미치는 요인들을 분석하여 과거 전력 사용량 데이터를 군집화한다. 군집화된 데이터를 대상으로 예측 일에 관련된 요인 데이터와 가장 유사한 데이터를 선정하고 이를 기반으로 예측 일의 전력 사용량을 예측한다. 예측 방법에 대한 성능 평가는 제안한 기법과 선행 연

구의 비교 실험을 실시하고, 그들간 예측치의 오차를 비교를 통하여 수행한다.

본 논문의 구성은 다음과 같다. 2장은 기존의 전력 수요를 예측한 방법들과 연구들을 살펴본다. 3장은 실제 전력 사용에 영향을 주는 요인을 분석하여 제안한 기법에 적용할 변수를 알아본다. 4장은 3장을 통해 살펴본 변수를 활용하여, 데이터 군집화를 통한 전력 사용량 예측 기법을 제안한다. 5장은 제안한 기법과 선행 연구와의 전력 사용량 예측 정확성을 통한 실험 및 평가를 수행한다. 마지막으로 6장에서는 결론을 맺는다.

2. 관련 연구

전력 사용량 예측 기법은 크게 장기, 중기, 단기로 나뉜다. 보편적으로 발전 시스템은 단기 예측 기법을 사용하는 경향이 크다. 기존의 단기 예측 기법은 시계열 예측 기법인 회귀 모형, 이동 평균법, 박스-젠킨스 모형의 예측 모형인 ARIMA 모형 등을 주로 사용하였다. 최근 연구 동향을 살펴보면, 우선 네 가지의 Neural Network 예측기에 각기 다른 시점의 과거 전력 사용량 데이터를 입력하여 최종 예측치를 얻는 단기 다중 예측 기법[2]이 제안되었다. 또한 요일 정보, 명절 등의 휴일의 패턴 특성을 고려한 전력 사용량 예측 기법[3] 및 계절 요인을 고려한 시계열 예측 기법[4, 5]과 기존 정보를 가지고 가중치를 주어 전력 사용량을 예측하는 기법[6] 등도 보고되었다. 하

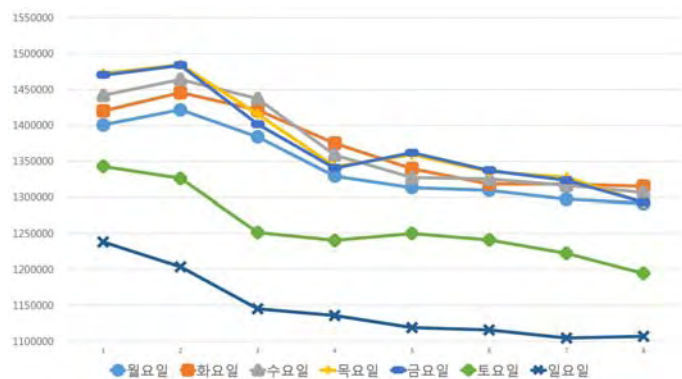
지만 이러한 선행 연구들은 예측 일에 대한 전력 사용량을 예측할 때, 여러 가지 복합적인 요인을 고려하지 않았다. 또한 전력 사용량은 기후 등 환경적 요인에 영향을 크게 받는다[6]. 본 연구는 이러한 복합적인 환경요인을 고려한 효과적인 전력 사용량 예측 기법을 제안한다.

3. 전력 사용 패턴에 영향을 주는 요인 분석

전력 사용량을 살펴보면 영향을 주는 다양한 요인들이 존재하며, 그로 인해 서로 다른 전력 사용 패턴이 존재한다. 그림 1은 2014년 3월 3일부터 4월 27일까지 8주간의 전국 전력 사용량을 요일별로 시각화 한 것이다.

1) 요일에 따른 전력사용 패턴

그림 1에서 일요일은 모든 기간에서 가장 낮은 전력 사용량을 보이고 있다. 다음으로 토요일이 낮은 전력 사용량을 보이며, 평일의 경우 월요일을 제외한 화, 수, 목, 금요일은 비슷한 전력 사용량을 보이고 있다. 하지만 월요일은 다른 평일에 비해 항상 낮은 전력 사용량을 보이고 있다. 이를 통해 요일마다 다른 전력 사용량 패턴이 보이고 있음을 확인할 수 있다.



(그림 1) 8주간의 전국 전력 사용량

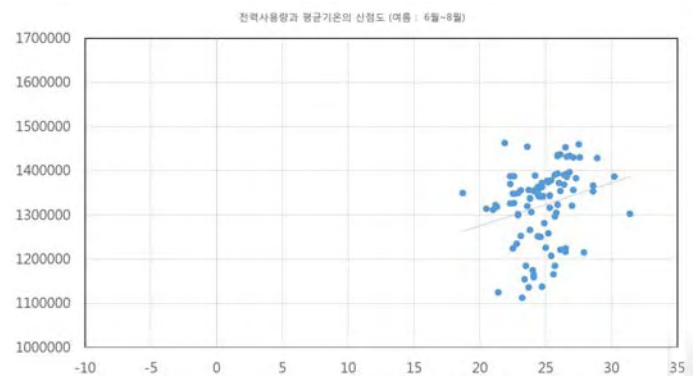
2) 기온과 전력 사용량 관계

그림 1에서는 각 요일마다 전력 사용 패턴이 비슷한 양상을 보이고 있다. 그러나 y축에 나타난 전력 사용량을 살펴보면, 각 주마다 높낮이가 다르다는 것을 알 수 있다. 이는 주마다 다른 기온 변화와 전력 사용량 변화의 상관관계가 크다는 것을 확인할 수 있다. 3월에서 4월로 넘어가는 시기는 겨울에서 봄으로 변하는 환절기로 기온의 변화가 큰 시점이다. 그림 1을 살펴보면, 환절기인 2주차에서 5주차 사이의 전력 사용량 변화가 크다는 것을 확인할 수 있다. 그림 2는 2014년 전국 전력사용량 데이터를 이용하여 여름과 겨울에 기온과 전력 사용량의 관계를 보여준다. x축은 기온을 나타내며, y축은 전력사용량을 나타낸다. 여름철에는 기온이 높아질수록 전력 사용량은 높아지고, 겨울철에는 기온이 낮아질수록 전력 사용량이 높아지는 것을 보이며, 기온이 전력 사용량에 큰 영향을 준다는 것을 알 수 있다.

4. 군집화를 통한 전력 사용량 예측 기법

본 장에서 데이터 군집화를 통하여 전력 사용량 예측의 정확성을 높일 수 있는 방안을 제안한다. 이를 위해 전력에 영향을 주는 환경 요인인 요일 정보 및 기온 정보를 수집하고, 유사한 환경을 가지는 전력 사용량 데이터를 군집화 한다. 이를 바탕으로, 예측 일의 환경 요인과 유사한 환경을 가지는 군집 데이터를 추출하고 이들의 전력 사용량을 예측에 활용한다. 제안한 기법의 시스템 구성은 그림 3과 같다.

(가) 여름철 기온과 전력 사용량의 산점도



(나) 겨울철 기온과 전력 사용량의 산점도



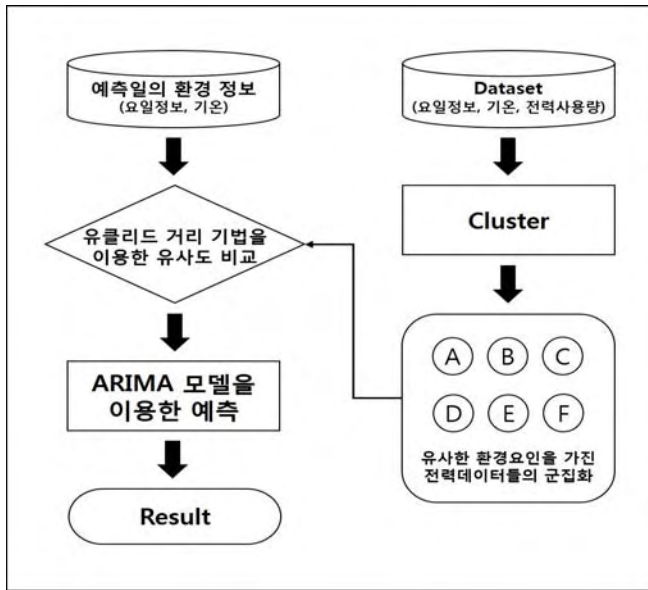
(그림 2) 여름, 겨울철 기온과 전력 사용량의 관계

4.1 환경 요인들을 고려한 데이터 군집화

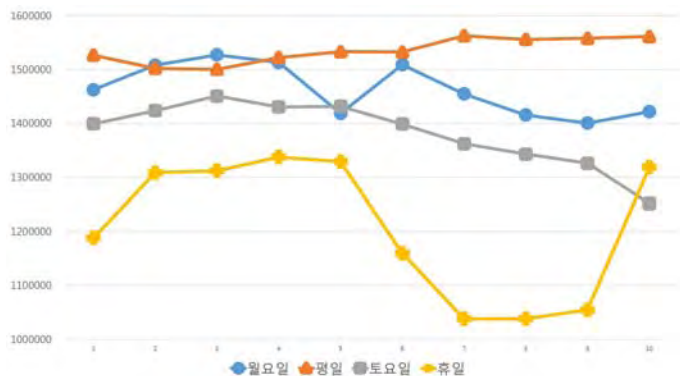
유사한 환경을 가질 경우, 전력 사용량 또한 유사하다는 것을 3장에서 제시하였다. 이러한 점을 가지고 유사한 환경을 가지는 전력 사용량 데이터를 군집화 하는 과정을 기술한다. 첫 번째, 비슷한 패턴을 가지는 요일별로 구분한다. 앞장에서 살펴본 것과 같이 월요일, 월요일을 제외한 평일, 토요일 그리고 일요일(휴일)로 구분된다.

그림 4는 요일을 네 가지로 구분한 전력 사용량 데이터를 시각화한 것이다. 휴일 및 일요일 그래프를 보면 유독 낮은 구간을 발견할 수 있다. 이 구간의 경우는 공휴일 및 명절과 같은 연휴이다. 연휴는 휴일이 연속되므로, 일요일 및 휴일보다 낮은 전력 사용량을 보인다. 이러한 점을 해

결하기 위해 표 1과 같이 요일 구분을 더 세분화하여 다섯 가지로 구분하였다.



(그림 3) 시스템 구성도



(그림 4) 4가지 요일로 구분한 전력 사용량

기온정보는 기상청에서 제공하는 공공 데이터의 해당 지역 일간평균 기온을 이용한다. 다음으로, 요일 및 기온 정보를 가지고 전력 데이터 군집화를 진행한다. 군집화 기법으로는 객체 간의 거리를 이용하여 유사도를 판단 후 군집화를 수행하는 K-Means 클러스터링 기법을 사용하여 유사한 환경을 가지는 전력 데이터들 간의 군집을 형성한다.

<표 1> 다섯 가지 요일 구분

요일 구분
월요일
화요일, 수요일, 목요일, 금요일
토요일
일요일
공휴일 및 연휴

4.2 환경 정보와 군집간의 유사도 비교 및 예측

예측일의 환경 정보는 요일 정보와 기상청에서 제공하는 동네 예보의 평균 기온 정보를 사용하였다. 유사도 비교 방법으로는 보편적인 거리 측정 기법의 하나인 유클리드 거리 기법을 적용하였다. 유클리드 거리 기법은 N차원에서 두 점 사이의 거리를 구하는 기법이며, 가장 직관적으로 유사도를 나타낸다.

$$\sqrt{(d_1 - f_1)^2 + \dots + (d_n - f_n)^2} = \sqrt{\sum_{i=1}^n (d_i - f_i)^2} \quad (1)$$

예측일의 환경 정보인 요일 정보와 기온 정보를 한 점으로 봤을 때, 각 군집의 중심점과의 거리를 식 (1)을 적용하여 계산한다. 가장 가까운 거리로 나타난 군집을 가장 유사한 군집으로 판단하고, 전력 사용량 예측에 사용한다.

예측 방법은 시계열 예측 모형 중 자기회귀 누적이동평균(Autoregressive Integrated Moving Average, ARIMA) 모형을 이용한다. ARIMA모형은 현재 가장 우수한 성능을 보이는 시계열 예측 기법으로 평가받고 있다. 분석 과정은 식별, 추정, 검증, 예측 등 네 단계로 이루어져 있다.

5. 실험 및 평가

본 장은 제안한 기법과 선행 연구에서 제안한 기법들을 대상으로 전력 사용량 예측을 수행한다. 실제 전력 사용량과 제안한 기법의 오차율을 비교하여 성능을 평가한다. 2011년 1월 1일부터 2013년 12월 31일간의 전력 사용량을 테스트 데이터 셋으로 사용하였다. 예측일 선정은 2014년도 임의의 20일을 선택하여 실험을 수행하였다. 임의로 예측일을 선정할 때는 한쪽으로 치우치지 않게 요일별로 다양하게 선정하였으며, 제안한 기법의 실험과 동일한 데이터를 이용하여 두 가지 비교 실험을 진행하였다. 첫 번째 비교 실험은 과거 동일 기간의 전력사용량 데이터를 적용 [7]하여 일간 예측에 적용하여 수행하였다. 두 번째 비교 실험은 계층적 군집방법을 이용한 기법[3]을 적용하여 비교 실험을 수행하였다.

평가 방법으로 평균 절대 백분율 오차(Mean Absolute Percentage Error, MAPE)를 사용하였다. 식 (2)는 MAPE 계산식을 나타낸다. 식 (2)의 n 은 총 횟수를 나타내며, X_d^{Actual} 은 실제 전력 사용량, $F_d^{Forecast}$ 는 예측된 전력 사용량을 말한다.

$$MAPE = \frac{1}{n} \sum_{d=1}^n \left| \frac{X_d^{Actual} - F_d^{Forecast}}{X_d^{Actual}} \right| \quad (2)$$

표 2에서 임의의 20일에 대한 예측일의 APE(Absolute Percentage Error)를 나타내었다. 표 3은 실험 전체에 대한 MAPE를 나타내었으며, 선행 연구에서 제안된 기법의

실험 결과와 본 논문에서 제안한 기법과의 오차율을 비교하였다.

<표 2> 임의의 20일에 대한 예측치 오차율

날짜	APE(%)		
	제안한 기법	비교논문 1	비교논문 2
2014-01-07	7.13%	0.43%	4.64%
2014-02-01	6.49%	18.07%	4.87%
2014-02-03	1.99%	12.22%	3.74%
2014-03-12	2.15%	2.68%	7.55%
2014-03-22	3.97%	5.07%	5.69%
2014-03-26	1.39%	6.92%	0.06%
2014-04-06	0.44%	9.01%	4.86%
2014-04-29	1.76%	0.77%	2.48%
2014-05-19	7.13%	1.73%	1.47%
2014-06-17	1%	7.01%	1.89%
2014-07-07	1.29%	6.58%	0.87%
2014-08-01	4.52%	7.06%	2.19%
2014-08-02	4.13%	2.77%	6.73%
2014-08-25	2.93%	0.09%	17.4%
2014-09-03	0.72%	17.2%	0.71%
2014-10-27	1.79%	1.48%	1.14%
2014-11-03	1.56%	4.49%	4.8%
2014-12-02	4.21%	7.13%	2.73%
2014-12-18	1.58%	7.41%	2.21%
2014-12-23	3.7%	0.32%	2.67%
2014-12-29	2.88%	1.28%	8.73%

표 3에서 제안한 기법과 비교 실험된 기법들을 비교하였을 경우, 제안한 기법이 비교 논문 1보다 52.88%, 비교 논문 2보다 35.35% 향상된 성능을 보였다. 이로써 제안한 전력 사용량 예측 기법의 정확성이 선행 연구보다 우수한 것으로 확인되었다.

<표 3> 임의의 20일에 대한 예측치의 MAPE

	제안한 기법	비교 논문 1	비교 논문 2
MAPE(%)	2.78%	5.90%	4.30%

6. 결론

본 연구는 정확한 일간 전력 사용량을 예측하기 위해 전력 사용량에 영향을 주는 환경요인들을 분석하였다. 또한, K-Means 클러스터링을 통해 유사한 환경을 가지는 전력 사용량 데이터를 군집화 하였다. 예측일과 유사한 환경을

가지는 전력 사용량 데이터를 적용하여, 기존의 예측 기법보다 정확성이 향상된 기법을 제안하였다. 제안한 기법을 적용하여 전력 사용량을 예측하였을 경우, 기존의 예측 기법들과 오차율을 비교한 결과 각각 52.88%, 35.35% 향상된 정확성을 보여 제안한 기법의 우수성을 확인하였다.

향후 시간대별 정확한 전력 사용량 예측에 관한 새로운 모델에 관해 연구할 계획이다. 또한, 전력 사용에 영향을 주는 다양한 요인과 패턴을 분석하여 더욱 정확성이 향상된 예측을 위한 연구를 진행할 계획이다.

7. Acknowledgements

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

참고문헌

- [1] 김성만, "마이크로그리드 기술의 적용과 운영사례," 대한전기협회 전기저널, sno. 455, pp. 37-44, 2014.11
- [2] P. P. K. Chan, W. C. Chen, W. W. Y. Ng and D. S. Yeung., "Multiple classifier system for short term load forecast of Microgrid," 2011 International Conference on Machine Learning and Cybernetics(ICMLC), Vol. 3, 2011.
- [3] H. M. Hwang, S. H. Lee, J. B. Park, Y. G. Park and S. Y. Son, "Load Forecasting using Hierarchical Clustering Method for Building," The Transactions of the Korean Institute of Electrical Engineers, Vol. 64, No. 1, pp. 41-47, 2015.
- [4] S. Y. Kim, H. W. Jung, J. D. Park, S. M. Baek, W. S. Kim, K. H. Chon and K. B. Song, "Weekly Maximum Electric Load Forecasting for 104 Weeks by Seasonal ARIMA Model," Journal of the Korean Institute of Illuminating and Electrical Installation Engineers, Vol. 28, No. 1, pp. 50-56, 2014.
- [5] H. W. Jung and K. B. Song. "Daily Maximum Electric Load Forecasting for the Next 4 Weeks for Power System Maintenance and Operation," The Transactions of the Korean Institute of Electrical Engineers, Vol. 63, No. 11, pp. 1497-1502, 2014.
- [6] K. B. Song, "Development of Short-Term Load Forecasting Algorithm Using Hourly Temperature," The Transactions of the Korean Institute of Electrical Engineers, Vol. 63, No. 4, pp. 451-454, 2014.
- [7] H. W. Jung, S. Y. Kim, K. B. Song, J. D. Park, S. M. Baek, W. S. Kim and K. H. Chon, "The Weekly Load Forecasting Using ARIMA Model," in Proceedings of The Korean Institute of Electrical Engineers, Jinju, 2013, pp. 91-92.