

시계열 패턴 군집분석을 기반으로 LSTM을 이용한 단기 전력 수요 예측

오민지, 김주하, 엄지성, 최은선, 김재성, 조완섭*

충북대학교 빅데이터학과, *충북대학교 경영정보학과

minji.oh1020@gmail.com, wldus1907@gmail.com, ejs0414@naver.com, tmxk147@gmail.com,

comkjsb@gmail.com, *wscho63@gmail.com

Short-Term Load Forecasting using Time Series Clustering and LSTM

Min Ji Oh, Joo Ha Kim, Ji Seong Eom, Eun Seon Choi, Jae Sung Kim, Wan Sup Cho*

Chungbuk National Univ. Department of Bigdata Science,

*Chungbuk National Univ. Department of Management Information

요약

고도화된 산업으로 인해 우리나라의 전력사용량은 증가하고 있어 전력사용량 예측에 대한 요구가 있으며, 정부의 스마트팩토리 추진 목표에 맞추어 설비가동률을 높이는 기업이 증가하고 있다. 스마트팩토리 운영에서 전기요금 감축과 에너지의 효율적 운용을 위한 전력 소비 패턴 관리 및 전력사용량 예측에 대한 중요성이 대두됨에 따라 본 연구에서는 건물의 특성과 요일에 따른 변화 및 시간대에 따른 특성을 시계열 분석에 반영하는 방안을 제안하였다. 형태 기반 거리척도인 DTW를 사용하여 요일과 시간대에 따라 시계열 데이터를 군집화한 뒤 LSTM 예측모델을 적용하여 군집화에 따른 전력 수요량 예측성능을 비교한 뒤 군집에 따른 전력 패턴의 차이를 분석하였다. K-means 군집화에 따라 군집의 개수를 2개로 설정한 뒤 군집의 특성을 반영한 LSTM모델의 예측력이 가장 뛰어났음을 확인하였다.

I. 서론

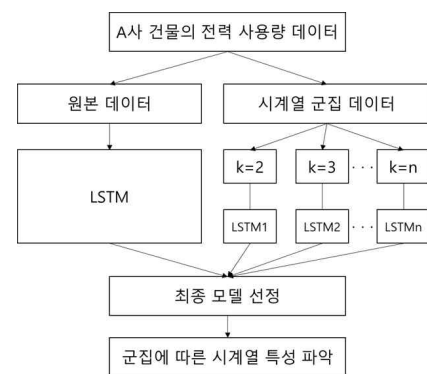
고도화된 산업으로 인해 우리나라의 전력사용량은 매년 증가하고 있다. 또한 정부의 스마트팩토리 추진 목표에 따라 국내 중소·중견기업에 스마트팩토리가 보급되면서 설비가동률을 높이는 기업이 증가하고 있다[1]. 스마트팩토리 운영에서 에너지 소비 감축은 주요 이슈로 떠오르고 있으며 건물의 효율적인 운용을 위한 에너지 소비 패턴 관리 및 예측에 대한 요구가 있다[2]. 건물의 에너지를 효율적으로 운용하려면 건물의 전력사용 패턴을 파악하고 이를 적절하게 분산시킬 필요가 있다. 특히 산업시설에 부과하는 전기요금은 전기 사용이 발생한 시간대와 요일을 구분하여 중부하, 최대부하, 경부하대로 나누어 요금을 산정하기 때문에 전력 사용 계획에 따른 전력 소비 분배는 생산비용을 낮추는 데 기여할 수 있다. 동일한 양의 전력을 사용하더라도 사용시간 및 사용패턴에 따라 전력요금이 달라지기 때문에 에너지의 효율적인 운용을 위해 시간적 패턴을 고려하여 전력 소비량을 예측해야하며 최대부하 시간대에 전력량을 줄이는 노력이 요구된다. 정확한 전력사용량을 예측하기 위해서는 건물의 개별 특성과 요일에 따른 변화, 시간대에 따른 특성을 반영해야한다. 하지만 이러한 요소들은 변화가 크기 때문에 정상성을 유지하기 어려운 특징이 있어 일반적인 통계적 시계열 분석에 반영하기 어렵다고 알려져 있다[3]. 본 연구에서는 요일과 시간대에 따른 시계열 특성을 반영하기 위해 형태 기반 거리척도인 DTW를 사용하여 데이터를 군집화한 뒤 LSTM 예측 모델을 적용하여 군집화에 따른 전력 수요량 예측 성능을 비교한 뒤 군집에 따른 전력 패턴의 차이를 분석한다.

II. 연구 방법

1. 분석프로세스

본 논문에서는 A사 건물의 전력 사용량 데이터를 활용하여 전력 사용량의 시계열 패턴을 기반으로 전력 수요 예측 분석을 수행하고 군집에 따

른 전력 사용량 시계열 특성을 파악하는 연구를 제안한다. 분석 구조도는 <그림1>과 같다. 원본데이터를 LSTM 모델에 적용한 전력 수요 예측과 시계열 데이터를 군집 개수(k)에 따라 나누어 LSTM 모델을 적용한 전력 수요 예측을 수행한 뒤 각 예측모델의 성능을 비교하여 최종 모델로 선정한다. 마지막으로 군집에 따른 시계열 데이터의 차이를 비교 및 분석한다.



<그림1> 분석 구조도

2.DTW

DTW는 Dynamic Time Warping의 줄임말로 형태를 기반으로 하는 탄력적 거리 측정 방식이다. 시계열 그래프 패턴 사이의 거리를 측정할 때 시간 축을 확장하거나 수축하는 방식으로 이를 왜곡하여 시계열 패턴 사이의 거리를 최적정렬 방법으로 찾는 알고리즘이다[4]. 주로 음성인식 분야에서 사용되며, 두 개의 시퀀스의 길이가 달라도 유사도를 측정할 수 있다는 장점이 있다. 시계열 군집분석에서는 거리 측정 방식으로 유클리디안(Euclidean)을 사용하지 않고 ETW를 사용한다. 개별 관측치 간의 동질성에 따라 거리를 측정하는 유클리디안 방식과 달리 DTW는 각 지점과 가장 유사한 위상의 점을 찾아 시계열 데이터의 시퀀스를 비교적 정확하

게 측정할 수 있다[5].

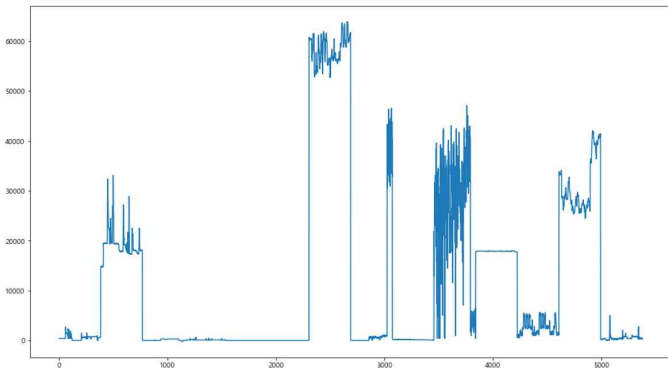
3.LSTM

LSTM은 Long Short Term Memory의 줄임말로 순환신경망(RNN)의 일종이다. 주기성을 가진 시계열 패턴을 예측하는데 사용되며 과거의 시계열 데이터와 현재의 데이터 조합으로 미래 데이터에 영향을 줄 수 있는 순환구조 형태를 지니고 있다[6]. 순환구조를 가진 특성으로 인해 음성인식, 문자, 시계열 예측 등의 분야에서 사용되고 있다.

III. 분석 결과

1. 데이터설명

본 논문에서 활용한 데이터는 A사 건물의 2015년 5월 3일부터 2015년 7월 28일까지 25일간 기록된 전력 사용 데이터이다. 센서에 의해 수집된 5분 단위 데이터를 1시간 단위로 합산하여 시간대별 시계열 데이터로 구성하였다. 전력 사용 데이터의 평일과 주말의 비율은 5:5로 구성되었다. 주7일에서 평일과 주말이 차지하는 비율이 5:2인 것을 고려하였을 때 연구에 사용된 데이터는 주말의 비율이 높은 것이라 판단할 수 있다. 또한 <그림2>를 통해 해당 기간의 부하유�효전력(W) 사용량의 시계열 분포를 확인하였다. 탐색적 데이터 분석을 통해 전력 사용량이 10,000 W 이상인 구간과 10,000 W 미만인 구간으로 나누어진다는 것을 알 수 있었다. 특정 기간에 따라 데이터의 시계열 분포가 달라진다는 점을 통해 A사 건물의 전력 사용량 데이터는 시계열 군집화 기법으로 전력 수요 예측을 하는 것이 적합할 것이라고 판단하였다.



<그림2> 전력 사용량의 시계열 분포

전력 수요 예측을 위한 시계열 군집화를 수행하기 위해 패턴 분류에 사용한 입력 변수는 <표1>과 같다. 연구에 사용한 데이터는 1년이 채 되지 않는 기간 동안 수집된 데이터이며 수집 기간이 5월에서 7월 사이이므로 '월' 변수와 '온도' 변수 데이터 포인트마다 차이가 두드러지지 않았기 때문에 패턴 분류의 입력 변수로 사용하지 않았다. <그림2>에서처럼 부하유�효 전력은 시간의 흐름에 따라 패턴을 보이므로 패턴 분류의 입력 변수로 사용하면서 하루를 기준으로 부하 유효 전력을 누적 계산한 부하누적 전력량(Wh)을 입력 변수에 추가하여 군집분류의 타당성을 높이고자 하였다. 또한 요일에 따른 패턴을 추적하기 위해 평일과 주말 여부를 확인할 수 있는 패턴 정보(Label)를 추가하였다.

입력 변수	설명
부하유�효전력 (W)	전원으로부터 공급되어 부하에서 유효하게 이용되는 전력
부하누적전력량 (Wh)	하루를 기준으로 한 누적 전력량 값 (매일 오전 0으로 리셋됨)
주말 여부	평일:0, 주말:1

<표1> 군집화에 사용한 패턴 입력 변수

전력 사용량 예측모델에 사용된 변수는 <표2>와 같다. 전력 사용량 예측에 있어 군집화 수행 유무가 예측 성능에 영향을 주는지 확인하기 위해 부하유�효전력과 군집화를 통해 부여한 군집번호를 변수로 활용하였다. 데이터가 수집된 기간이 짧아 데이터의 총 레코드가 8,064개 행인 점을 감안하여 과거 3시간 동안의 전력 사용량 데이터와 군집번호를 바탕으로 1시간 이후의 전력 사용량을 예측하는 방법을 사용하였다.

변수	설명	비고
전력사용량	1시간 이후의 전력 사용량	종속변수
전력사용량 (t, t-1, t-2, t-3)	현재 및 과거 3시간의 전력 사용량	독립변수
군집번호	군집화 결과 번호	

<표2> 전력 사용량 예측 모델의 변수

2. 성능평가지표

본 논문에서 예측모델의 성능 평가지표로 사용한 것은 MAE, RMSE, R2이다. MAE는 식<1>과 같이 실제 값과 예측 값의 차이를 절대값으로 변환해 평균화한 것으로서 오차의 크기가 그대로 반영된다는 장점이 있다. RMSE는 식<2>와 같이 실제값과 예측값의 차이를 제곱하여 평균을 계산한 MSE의 제곱근을 계산한 것이다. MSE는 오류의 제곱을 구하므로 실제 오류 평균보다 더 커지는 문제가 있어 이를 해결하기 위해 RMSE를 사용하였다. R2는 결정계수라고 불리며 추정한 모형이 주어진 자료에 적합한 정도를 계산하는 척도이다. 식<3>과 같이 R2는 종속변수 y의 분산 중에서 적합된 모형이 설명가능한 부분의 비율을 의미하므로 적합된 모형의 설명력을 측정할 수 있다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{T}} \quad (2)$$

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \quad (3)$$

3. 실험결과

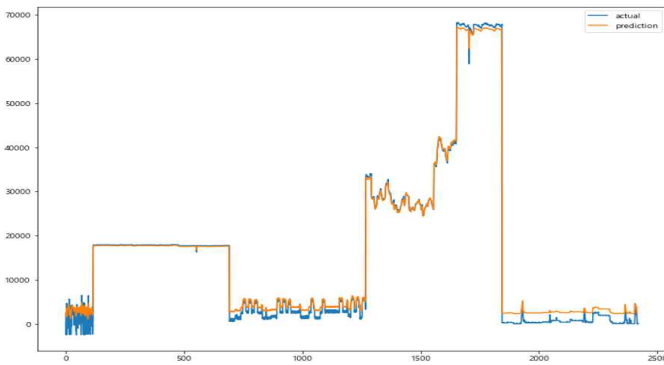
군집화의 개수에 따른 예측 성능을 비교하기 위해 군집화를 진행하지 않은 데이터와 k=2, 3, 4, 5로 설정하여 군집화를 수행한 데이터의 성능을 비교하였다. 시계열 군집분석은 DTW를 이용한 Time Series K-Means 기법을 사용하였으며, 전력 수요량 예측 알고리즘은 LSTM을 사용하였다. 군집의 개수를 2에서 5 사이로 설정하여 군집 개수에 따른 예측 성능을 비교하였다. 예측 성능의 개선 여부를 확인하기 위해 군집화를 사용하지 않은 데이터 학습 결과를 기준으로 사용하였다. <표3>과 같이 군집 분석을 수행한 예측모델은 모든 k에 대해 RMSE는 기준점보다 낮고, R2는 기준점보다 높다는 것을 확인하여 군집분석을 통한 시계열 예측이 효

과적임을 입증하였다. 가장 예측성능이 뛰어난 모델은 군집을 2개로 나누고 군집번호를 독립변수로 반영하여 LSTM에 적용한 것으로 드러났다.

	기준	k=2	k=3	k=4	k=5
MAE	1187.13	1050.24	1479.08	1694.90	1658.92
RMSE	3005.30	1560.39	2379.09	2792.62	2976.78
R2	0.94	0.99	0.98	0.97	0.97

<표3> 군집화에 따른 전력 사용량 예측 성능 비교

<그림3>은 군집의 개수(k)를 2로 설정한 뒤 평가 데이터로 실제값과 예측값의 일치여부를 시각적으로 확인한 것이다. 시간의 흐름에 따른 전력 사용량의 변화를 정확하게 잡아내었다는 것을 알 수 있다.



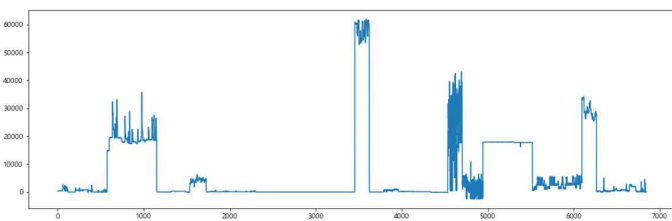
<그림3> 전력 수요량 실제값과 예측값 비교

두 개의 군집으로 나누어 어떤 차이를 보이고 있는지 알아보기 위해 군집1과 군집2의 전력 사용량에 대한 기술통계량을 <표3>과 같이 확인하였다. 두 군집은 평균 전력사용량과 전력사용량 중앙값에서 큰 차이를 보였다. 그러므로 시간에 따른 전력 사용량 패턴에서 구분되는 차이점이 발견되었을 때 적절한 패턴 군집화 방법을 적용하는 것이 전력 수요량을 정확히 예측하는 데 기여할 수 있다고 판단할 수 있다.

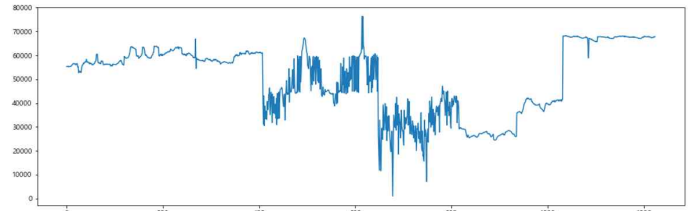
	군집1	군집2
비율	전체의 84%	전체의 16%
평균	6469 W	50047 W
중앙값	327 W	56218 W
표준편차	11929	14395

<표3> 군집에 따른 기술통계량 비교

군집1에 속하는 데이터의 시계열 그래프 <그림4>와 군집2에 속하는 데이터의 시계열 그래프 <그림5>를 비교하였다. <그림4>는 전력 사용량이 10,000 W를 넘지 않는 데이터가 다수인 반면 <그림5>는 전력 사용량이 10,000 W 이상인 데이터가 분포하고 있음을 확인할 수 있다.



<그림4> 군집1의 전력 사용량 시계열 그래프



<그림5> 군집2의 전력 사용량 시계열 그래프

IV. 결론

본 논문에서는 요일과 시간대에 따른 시계열 특성을 반영하기 위해 형태 기반 거리척도인 DTW를 사용하여 데이터를 군집화한 뒤 LSTM 예측 모델을 적용하여 군집화에 따른 전력 수요량 예측 성능을 비교한 뒤 군집에 따른 전력 패턴의 차이를 분석하였다. 원본데이터를 LSTM 모델에 적용한 전력 수요 예측 모델과 군집 개수(k)에 따라 시계열 데이터를 나누어 LSTM 모델을 적용한 전력 수요 예측 모델의 성능을 비교한 결과, 군집을 2개로 나누어 LSTM을 수행한 모델의 예측 성능이 가장 우수했다. 두 군집의 기술통계량을 비교한 결과 평균 전력 사용량과 전력사용량의 중앙값에서 유의한 차이가 있음을 확인하였다. 이처럼 시계열 패턴 사이에 차이점이 발견되었을 때 적절한 패턴 분류 방법을 적용하여 학습 모델을 생성한다면 개별 건물 특성과 시간적 특성을 반영할 수 있어 좀더 정확한 전력 수요량 예측을 가능하게 할 수 있다고 판단할 수 있다. 다만 연구에 사용된 데이터의 수집 기간이 1년이 되지 않는다는 점에서 다양한 계절적 특성을 반영할 수 없었다는 것이 한계점으로 남아 있다. 향후 데이터 수집 기간을 늘려 계절적 특성을 반영할 수 있는 시계열 패턴 분석을 수행하여 연구의 확장성을 이어나갈 것이다.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2018-0-01396) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation)

참 고 문 헌

- [1] 정종필, 신광섭(2020). 융합연구리뷰. 융합연구정책센터. Vol6.
- [2] 박재연, 허태욱, 고정길(2021). 한국통신학회 학술대회논문집, p132-133.
- [3] Yang, L., & Yang, H(2019). Analysis of different neural networks and a new architecture for short-term load forecasting. Energies, 12(8), 1433.
- [4] 이에담, 임재현, 최동구(2020). 형태 기반 군집화와 랜덤포레스트를 이용한 개별 건물의 단기 전력 수요 예측. 경영과학 37(4), 21-31.
- [5] Paparrizos, J., & Gravano, L.(2015). k-shape: Efficient and accurate clustering of time series. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data (pp. 1855-1870).
- [6] 김진석, 김기일(2019). 전력 수요관리를 위한 전력사용량 패턴예측에 관한 연구. 한국통신학회 학술대회논문집.