

# 전력 데이터의 특징 추출 및 XGBoost 를 이용한 숙박 업소 재실 여부 판단

김에덴\*, 고석갑\*, 손승철\*, 이형옥\*, 이병탁\*  
\*한국전자통신연구원

{kimed93, softgear, sson, oklee, bytelee}@etri.re.kr

## Determination presence of people in accommodation using feature extraction and XGBoost method of energy data

Eden Kim\*, Seok-Gap Ko\*, Seung-Chul Son\*, Hyung-Ok Lee\*, Byung-Tak Lee\*  
\*Electronics and Tele-communications Research Institute (ETRI)

### 요 약

스마트미터의 기술 발달과 보급으로 인해 전력데이터의 수집이 보다 수월 해짐에 따라 각 시스템에 효율적인 맞춤 서비스 제공을 위한 전력 데이터 분석 기술에 관한 다양한 연구가 활발하게 진행되고 있다. 관련하여 본 논문에서는 숙박업소의 각 방마다 전력소비량을 측정 및 수집하여 전력소비패턴을 분석하고 특징 추출 및 XGBoost 를 이용한 머신러닝 분석방법으로 각 방의 사람 재실 여부를 판별하는 방법을 소개한다. 이와 같은 연구를 통해 추후 숙박업소 혹은 숙박업소를 이용하는 소비자들의 맞춤 서비스 제공에 응용 및 적용 할 수 있다.

### 1. 서론

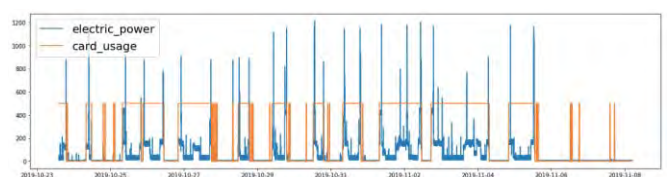
최근 에너지 분야에서 전력의 효율적인 사용과 관리를 위해 스마트 그리드 기술이 활성화 되면서 스마트미터의 보급이 늘어나고 있고, 이와 함께 지능형 시스템을 위한 기술 개발이 다양하게 이루어지고 있다 [1]. 특히, 보급된 스마트미터를 이용하여 에너지 전력 사용량을 실시간으로 측정하고 정보를 주고받는 기술들이 발달하면서 학교, 회사, 관공서 등 각 사용처에 맞춤 서비스 제공을 위해 전력 데이터 분석을 비롯한 많은 연구들이 진행되고 있다 [2]. 이에 일환으로, 본 논문에서는 전력 데이터를 이용한 숙박업소에서 사람의 재실 여부 분석 기술을 소개한다.

본 연구에서는 숙박업소의 각 방에 해당하는 전력 소비 데이터와 카드키 on/off 정보를 수집한다. 카드키 정보는 사람의 재실 여부를 파악하는 정보로 활용되며 기존 과거 전력 데이터와 카드키 정보를 통하여 머신러닝 모델 학습을 하고, 추후 전력 데이터만을 이용하여 재실 유무를 높은 정확도로 판별 가능함에 관한 실험을 진행하였다. 가장 간단하게는 전력 신호와 카드키 정보를 매칭하여 임계 값 설정을 통한 판별 방법 혹은 원시데이터를 이용한 머신러닝 분류

방법 등을 고안해 볼 수 있다. 하지만 보다 높은 정확도를 보이는 분석 알고리즘을 만들기 위해서 본 연구에서는, 전력 데이터는 시계열 특징을 가지는 것을 고려하여 시계열 데이터가 가질 수 있는 특징들을 생성 및 추출하였고, 원시데이터를 바로 사용하는 것이 아닌 새로 추출한 특징 벡터와 카드키 정보의 관계 매칭을 통해 XGBoost 방법을 적용함으로써 기존 단순한 판별 기술보다 더 우수한 결과를 나타내는 분석 알고리즘을 개발 하였다.

### 2. 데이터 수집 및 전처리

본 연구를 진행하기 위해 센서를 통하여 1 초단위의 전력 데이터와 카드키 정보를 실시간으로 수집하여 데이터베이스에 저장하였다. 수집한 전력 데이터와 카드키 정보의 관계는 다음 그림 1 과 같이 서로 연관성이 있음을 보여준다.

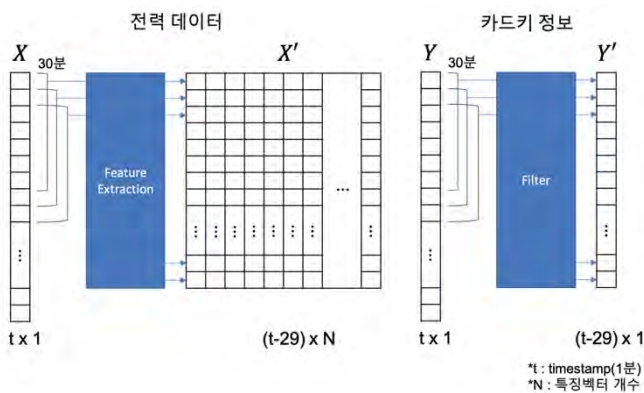


(그림 1) 전력데이터와 카드키 정보의 관계

각 데이터는 초단위로 수집되었지만, 재실 유무를 초 단위까지 고려하여 판별할 필요는 없기 때문에, 모든 데이터를 1분 단위로 평균을 통한 다운샘플링을 진행하였다. 전력 데이터는 방마다 기저 전력이 다르고 피크 전력량이 다르기 때문에 min-max 정규화를 통하여 모든 방의 전력들의 스케일을 조정하였다. 또한, 카드키 정보 중에는 1분 단위 기준으로 30분 이내 시간 동안 카드키가 on이 되었다 다시 off가 되는 경우가 간혹 존재하는데, 이의 경우는 실제로 손님이 대실 혹은 숙박의 목적으로 머무르고 있는 상황이 아니라 방 청소 혹은 잠시 물건만 놔두고 다시 외출하는 등 숙박업소에 재실 하는 경우와 다른 상황일 확률이 높아 30분 이내 기준으로 on되어있는 값들은 모두 off로 간주하도록 전처리를 진행하였다.

**3. 특징 추출 및 XGBoost 를 이용한 분석**

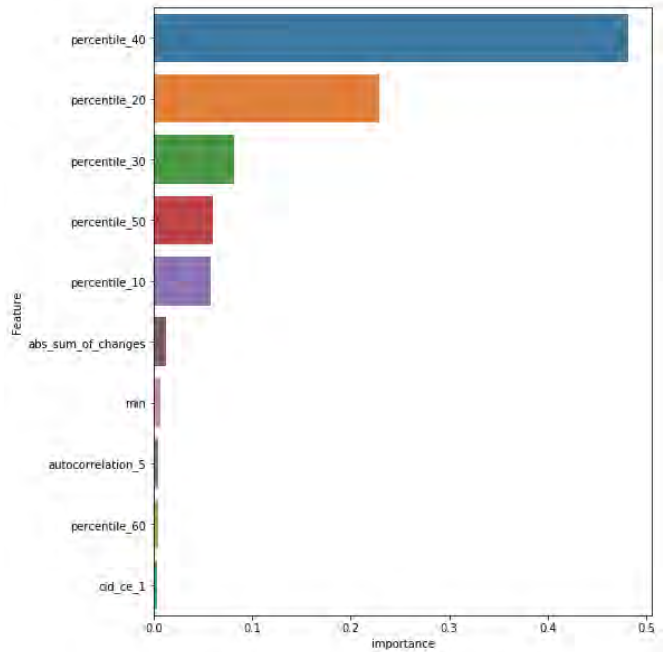
본 3 장에서는 전처리 된 전력 데이터와 카드키 정보를 통하여 시계열 특성을 반영한 특징 벡터들을 먼저 추출한다. 시계열 데이터의 특징 추출 방법에 관하여 다양한 연구들이 진행 되었는데, 그 중에 파이썬 패키지중 하나인 tsfresh 를 통한 특징 추출의 방법들에 대하여 참고하였다 [3]. 2 장에서 30분을 기준으로 전처리를 진행하였던 바와 같이 본 연구에서 사용하는 데이터는 30분을 기준으로 데이터의 의미를 부여하여 30분동안의 시계열 전력 데이터 세트를 통하여 평균, 최댓값, 최솟값 등을 포함한 58 가지의 시계열 특징 벡터들을 추출하였다. 카드키 정보의 경우는 30분을 기준으로 필터링 하여 30분동안 on인 경우가 off인 경우보다 많으면 on, off인 경우가 on인 경우보다 많으면 off로 카드키 정보를 재구성 하였다. 그림 2는 전력데이터의 특징 추출하는 과정과 카드키 정보를 필터링 하는 과정을 보여준다.



(그림 2) 데이터 셋의 특징 추출 및 필터링 과정

다음 진행으로는 특징 추출된  $X'$ 와 필터링 된  $Y'$  데이터 셋을 가지고 XGBoost 모델을 사용하여 학습

하였다. XGBoost 는 캐글 커뮤니티에서 각광 받았던 트리 기반 부스팅 방법으로 빠르고 유연하게 사용 가능하고 보다 직관적이다 [4].



(그림 3) 모델의 상위 10 가지 Feature Importance

위 그림 3 은  $X'$ 와  $Y'$ 를 XGBoost 로 학습한 모델의 특징 중요도를 나타낸다. 선택한 58 가지 시계열 특징들 중 가장 중요한 상위 10 가지 특징들을 살펴보면 대부분 Percentile 특징들이 재실 여부를 판별하는데 중요한 정보를 가지고 있음을 알 수 있다.

**4. 실험 결과 및 타당성 검증**

3 장에 기술한 특징 추출 및 XGBoost 를 통한 분석 결과 검증을 위해 아래의 표 1 과 같이 실험 데이터를 구성하였다.

<표 1> 실험 데이터 셋 구성

호	날짜	
	2019.10.23 ~ 2019.11.15	2020.01.23 ~ 2020.02.12
101	A1	A2
102	B1	B2
103	C1	C2

실험에 사용된 데이터는 한 숙박업소의 101 호, 102 호, 103 호 방 세 곳의 전력 및 카드키 정보를 수집한 것이고, 각각은 19년 10월 23일부터 11월 15일까지의 데이터와 20년 1월 23일부터 20년 2월 12일까지의 데이터로 나누어 각각 A1, A2, B1, B2, C1, C2 로 표기하였다.

다음과 같은 데이터 셋을 이용하여, 첫번째로는 같

은 방의 데이터를 사용한 트레이닝 및 테스트를 하는 방식, 두번째로는 두 방의 데이터를 트레이닝, 두 방 이외의 나머지 다른 방을 테스트 하는 방식으로 진행하여 기존 전력 원시데이터를 이용한 XGBoost 분석과 제안된 특징 추출을 통한 XGBoost 분석을 비교 실험하였다. 아래의 표 2 는 실험 데이터 셋에 따른 분석 결과 비교를 나타낸다.

<표 2> 실험 데이터 셋에 따른 판별 결과 비교

	실험 데이터 셋	판별 정확도	
		기본 방법	제안 방법
1	A1 트레이닝 A2 테스트	97.28%	99.78%
	B1 트레이닝 B2 테스트	94.04%	98.22%
	C1 트레이닝 C2 테스트	81.45%	98.21%
2	A1, B1 트레이닝 C2 테스트	82.27%	98.22%
	B1, C1 트레이닝 A2 테스트	83.31%	99.88%
	C1, A1 트레이닝 B2 테스트	73.62%	93.86%

전반적으로 같은 방의 데이터를 트레이닝 하고 테스트 하는 경우는 정확도가 대부분 90% 이상으로 높게 나왔으나, 트레이닝과 테스트 하는 방의 데이터가 다른 경우는 비교적 낮은 정확도를 보였다. 그 중에서도 원시데이터를 이용한 기본 분석 방법을 취했을 때는 70~80%의 낮은 정확도를 보였으나, 제안된 방법은 더 높은 정확도를 보여 이 논문에서 제안한 방법이 더 우수한 성능을 보이는 것을 확인하였고 두번째와 같은 실험을 통하여 트레이닝 셋에 포함되지 않은 다른 방을 테스트 하는 경우에도 여전히 90% 이상의 높은 정확도를 보여 실제 학습에 사용되지 않은 방의 경우도 전력데이터만 보유하고 있으면 꽤 높은 정확도로 사람의 재실 여부를 판별 할 수 있다.

**5. 결론**

스마트미터로부터 측정되는 전력데이터를 통하여 분석함으로써 여러 서비스들을 위한 정보를 제공 할 수 있다. 본문에서는, 특히 숙박 업소에 측정되는 에너지 전력 데이터를 통하여 사람 재실 여부를 판별하는 분석 방법에 대하여 소개하였다. 기본적인 방법으로 원시데이터를 머신러닝 적용하여 판별하는 것보다 본 논문에서 제시된 시계열 특징을 지닌 전력데이터의 특징 벡터를 뽑아 판별 분석하였을 때 더 뛰어난 정확도를 보임을 실험을 통하여 검증하였다. 본 연구에

나아가서는 본 논문에서 선택했던 58가지의 특징벡터들을 최적화 하는 방법, 결과의 후처리 등을 통하여 정확성을 더 높일 수 있을 것으로 보이며, 이와 같은 재실 판별의 정확도가 높아지면 숙박업소와 숙박업소를 이용하는 사람들을 위한 서비스 응용에 적용 가능할 것으로 보인다.

**감사의 글**

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 20181210301570)

**참고문헌**

[1] 이일우, et al. “스마트 그리드 기술 동향.” *한국통신학회지 (정보와통신)*, vol. 26, no.9, pp.24-33, 2009.  
 [2] Alahakoon, Daminda, and Xinghuo Yu. “Smart electricity meter data intelligence for future energy systems: A survey.” *IEEE Transactions on Industrial Informatics*, vol.12, no.1, pp.425-436, 2015  
 [3] Christ, Maximilian, et al. “Time series feature extraction on basis of scalable hypothesis tests (tsfresh-a python package).” *Neurocomputing*, vol.307, pp.72-77, 2018.  
 [4] Chen, Tianqi, and Carlos Guestrin. “Xgboost: A scalable tree boosting system.” *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, San Francisco, USA, 2016, pp.785-794.