

설명 가능한 인공지능 매개 에너지 수요 예측 및 효율성 연구

신지혜, 김윤재, 이수진, 문현준

세종대학교

hmoon@sejong.ac.kr

Study on Heat Energy Consumption Forecast and Efficiency Mediated Explainable Artificial Intelligence (XAI)

Jihye Shin, Yunjae Kim, Sujin Lee, Hyeonjoon Moon

Sejong University

요 약

최근 전세계의 탄소중립 요구에 따른 에너지 효율 증대를 통한 에너지 절감을 위한 효율성 관련 연구가 확대되고 있다. 방송과 미디어 분야에는 에너지 효율이 더욱 시급하다. 이에 본 연구에서는 효율적인 에너지 시스템 구축을 위해 난방 에너지 시계열 데이터를 기반으로 한 수요 예측 모델을 선정하고, 설명하는 인공지능 모델을 도입하여 수요 예측에 영향을 미치는 원인을 파악하는 프레임워크를 제안한다.

1. 서론

한국과 유럽 등 전세계가 탄소중립을 추구함에 따라 에너지 효율 증대를 위한 연구와 시장은 확대되고 있다. 탄소중립을 실현하기 위해 에너지 절감은 반드시 실현되어야 할 목표이다. 에너지 절감의 방향은 크게 하드웨어 개선과 소프트웨어 구축으로 나눌 수 있는데 본 연구에서는 에너지와 연관성 있는 데이터 구축 및 모델 개발을 통해 에너지 효율을 높이고 설명 가능한 인공지능 모델(XAI)을 이용해 에너지 수요 예측에 필요한 변수 선정 및 에너지 효율을 높일 수 있는 설명력을 강화하는 모델을 구축한다[1]. 최근 들어 인공지능 기술 영역이 스마트 빌딩 구축에 영향을 미치고 있는 것은 탄소중립 정책과 무관하지 않다. 스마트 빌딩 개념에 대한 이해가 IoT 시스템에 집중되어 있지만 에너지가 반드시 필요한 빌딩 구조에서 에너지 절감에 대한 요구는 높아지고 있는 현실이다. 디지털 방송과 정보 전달 수단의 다양한 미디어도 에너지 절감에 앞장서야 하는 위치에 있다. 이렇듯 일상에서부터 빅데이터를 다루는 모든 분야를 포함해 에너지 수요 예측은 중요한 문제 해결 주제이며 동시에 수요 예측에 대한 설명력을 강화한다.

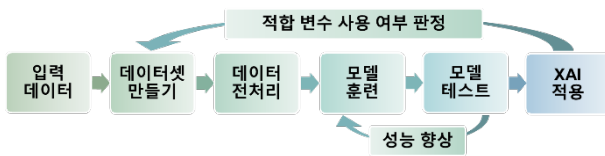
본 연구는 10 년동안 수집한 난방 에너지 시계열 데이터를 기반으로 설계되었으며 시계열 데이터의 특성을 고려하여 데이터를 분석하고 처리하였다. 우리나라에서 수집하는 시계열 데이터는 시간의 흐름에 따른 변화가 규칙적이나 변화를 기반으로 한다. 따라서 학습과 검증 및 테스트를 위한 데이터셋의 분할 및 구성 등 데이터 처리 그리고 다양한 머신러닝 모델을 적용하여 정확도를 높여 신뢰도 있는 모델을 구축하는데 연구력을 집중했다. 열에너지의 효율적인 관리 시스템을 위해서는 열에너지의 수요를 예측하고 이러한 수요가 나타나게 된 원인을 XAI 를 통해 설명한다.

2. 방법론

우리는 청주지역 난방공사가 제공한 87,672 개의 시간별 지역난방열소비량 데이터셋을 활용하여 데이터 전처리 및 모델 분석 그리고 프레임워크를 구성했다. 구성 순서는 다음과 같다. 먼저, 입력 데이터로 활용할 변수를 선정한 후, 일시를 기준으로 한 시계열 데이터셋으로 만든다. 그리고 데이터셋의 전처리를

위해 결측치 제거 및 정규화를 진행한다. 정확한 수요 예측을 위해 모델 훈련과 테스트 방법을 다각화하고 각 모델에 XAI 를 적용하여 적합한 변수 선정 및 그 이유에 대한 설명을 매개체로 하여 에너지 효율을 높이는 시각적 도구로 활용한다.

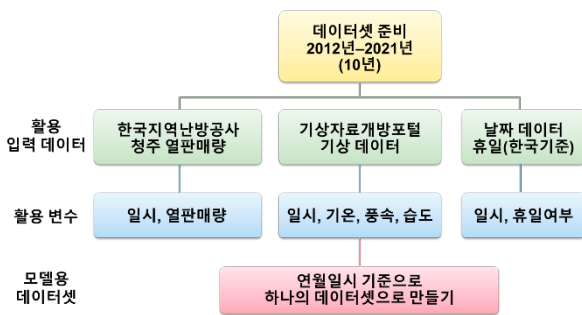
본 연구에서 사용한 모델은 LightGBM, XGBoost, AdaBoost, MLP, 그리고 LSTM 으로 각 모델의 성능 비교 및 평가를 통해 정확성이 보장된 수요 예측 모델을 선정한다. 마지막으로 설명을 위한 XAI 모델로 SHAP 을 활용하여 모델에 영향을 미치는 변수의 중요도를 파악하고 변수 선정에 활용하기 위한 방안에 대해서 검토한다.



<그림 1> 방법론

3. 데이터셋 구성 및 전처리

열 에너지 수요 예측을 위해 필요한 변수로 기상 데이터와 휴일 정보 그리고 월과 시간 정보를 선정한다. 날씨와 관련된 기상 데이터는 열수요와 밀접한 관련이 있다[2]. 또한, 계절적 특성은 시계열 데이터의 중요한 변수가 될 수 있다. 따라서 월 정보와 시간에 따라 날씨가 달라진다는 점을 활용할 수 있도록 시간 정보를 변수에 포함한다[3]. 마지막으로 특정 건물, 일하는 사람의 수 등이 영향을 미칠 확률을 고려하기 위해 휴일 정보를 포함한다.



<그림 2> 데이터셋 구성

본 연구에서는 10 년치 데이터의 활용을 위해 2012 년부터 2021 년까지의 기간에 수집된 데이터를 사용하며 이 중 2012 년부터 2020 년까지(총 9 년)는 훈련을 위해 사용하고, 2021년(총 1년)은 테스트 데이터셋으로 사용한다.

<그림 2>와 같이 총 3 가지의 데이터셋을 사용한다. 첫째로 공공데이터포털과 한국지역난방공사에서 제공하는 시간별 청주 지점의 열판매량(열공급량)이다. 데이터 구성은 1 시간 간격으로 나누어져 있으며 변수는 일시와 열판매량이다. 둘째는 기상자료개방포털에서 제공하는 청주 지역의 기상데이터이다.

동일하게 1 시간 간격으로 데이터가 제공되며 변수는 일시, 기온, 습도, 풍속이다. 마지막으로 공공데이터포털 또는 달력을 통해 데이터를 제공받아 휴일인 경우 1, 평일인 경우 0 으로 라벨링한다. 휴일은 청주가 있는 한국을 기준으로 토요일, 일요일, 공휴일, 대체공휴일을 모두 휴일로 정의한다. 일시는 모두 동일하게 연, 월, 일, 시간 정보를 담고 있는 datetime64[ns] 형식이다. 그리고 기온, 습도, 풍속은 소수점 자리 표기가 있는 float64 형식이고, 휴일과 열판매량은 정수로만 표기된 int 형식이다. 일시 데이터를 기준으로 데이터를 병합한 후, 일시 데이터에서 월 정보와 시간 정보를 추출하여 컬럼에 추가한다.

데이터셋 전처리를 위해 변수명을 모두 영어로 명기한다. 일시는 DateTime, 기온은 temp, 습도는 humidity, 풍속은 windspeed, 휴일은 holiday, 월은 month, 시간은 hour, 열판매량(열공급량)은 Amount 로 표기하고 데이터셋의 결측치 여부를 확인한다. 본 데이터셋에서는 기상 데이터의 일부가 결측이 되어 이전 시간대의 데이터로 대체하는 방법을 통해 결측치를 처리하였다. 마지막으로 각 column 은 모두 다른 단위를 가지고 있기 때문에 정규화를 통해 0 과 1 사이의 값들로 만들기 위해 MinMaxScaler 를 사용했다.

4. 시계열 예측 모델에서의 입력 및 출력 변수

구성한 데이터셋에서 입력 변수인 기온, 습도, 풍속, 휴일, 월, 시간과 타겟 출력 변수인 열판매량을 분류해주는 작업을 한다. 수요 예측에 있어서 입력 변수와 출력 변수의 시점을 정하는 기준을 두 가지 방법으로 나누어 예측일의 목적에 따라 방식을 다르게 적용한다. 첫째는 <그림 3(a)>에서와 같이 동일한 시간의 입력 변수와 출력 변수를 쌍으로 훈련하여 알고 싶은 날에 대해 예측된 기상 정보와 시간, 휴일 등 알고 싶은 날의 정보를 입력으로 사용하여 미래를 예측하는 방법이다.



<그림 3> 입/출력 변수 분리 방법 예시

다른 방식은 <그림 3(b)>와 같이 현재 시간의 입력 변수와 미래 시간의 출력 변수를 쌍으로 훈련하여 현재 기상 정보를 입력으로 사용하여 지정된 미래 정보를 예측하는 것이다. 이 방식에서는 현재와 미래 시간의 간격에 따라서도 다르게 볼 수 있는데, 본 연구에서는 1 시간 간격과 24 시간 간격으로 하는 경우에 대해서 실험했다.

<표 1>은 동일한 환경에서 입력 변수와 출력 변수를 분리하는

방식을 다르게 하여 LightGBM과 LSTM에 적용하고 각각 비교한 결과이다.

<표 1> 입/출력 변수 분리 방식을 모델에 적용한 성능 비교 결과

LightGBM 모델	R2 score	RMSE	MAPE
방법 1	0.923	249.663	0.172
방법 2 [1시간 간격]	0.922	252.545	0.172
방법 2 [24시간 간격]	0.895	334.811	0.184
LSTM 모델	R2 score	RMSE	MAPE
방법 1	0.933	216.774	0.166
방법 2 [1시간 간격]	0.925	240.593	0.167
방법 2 [24시간 간격]	0.887	359.362	0.185

결과값은 모두 소수점 넷째 자리에서 반올림하여 셋째 자리까지 표기하였다. 두 모델에서 방법 1의 성능이 가장 좋은 것을 확인했는데, 이는 방법 2의 1시간 단위와 비교하였을 때보다 방법 2의 24시간 단위(일 단위)와 비교하였을 때 더 큰 차이를 보였다. 이에 따라 방법 1의 입력 및 출력 변수 분리 방식을 활용하기로 결정했다.

5. 모델 평가 및 XAI 적용

열에너지 수요 예측을 위한 모델로 사용한 LightGBM Regressor, XGBoost Regressor, AdaBoost Regressor, MLP Regressor, LSTM, 5가지 모델에 대해 간략하게 설명을 하자면 다음과 같다. LightGBM은 GOSS와 EFB를 적용하여 GBDT를 새롭게 구현한 모델로 기존의 GBDT보다 훈련 과정을 최대 20배 이상 빠른 속도로 거의 동일한 정확도를 달성하게 하는 모델이다[4]. XGBoost는 확장 가능한 end-to-end tree boosting system으로 빠르고 정확한 모델이다[5]. AdaBoost는 이전 라운드의 가중치를 반영하여 다음 라운드에 반영하는 방식으로 과적합되는 경향을 줄여주는 특징이 있는 모델이다[6]. MLP는 다층 퍼셉트론(Multi-Layer Perceptron)의 준말로 XOR 문제를 해결하기 위해 등장한 모델로 입력층과 은닉층 그리고 출력층으로 구성된 것이 특징인 뉴럴 네트워크이다[7]. LSTM은 RNN의 단점을 보완한 모델로 일종의 장단기 메모리(Long Short-Term Memory)의 줄임말이며, 은닉층의 메모리 셀에 입력 게이트, 망각 게이트, 출력 게이트를 추가하여 불필요한 기억을 지우고, 기억해야 할 것을 정하는 모델이다[8]. 수요 예측을 위해 5개의 모델을 동일한 데이터셋을 이용하여 훈련 후 테스트 데이터에 대한 평가 지표가 좋은 모델을 찾는다.

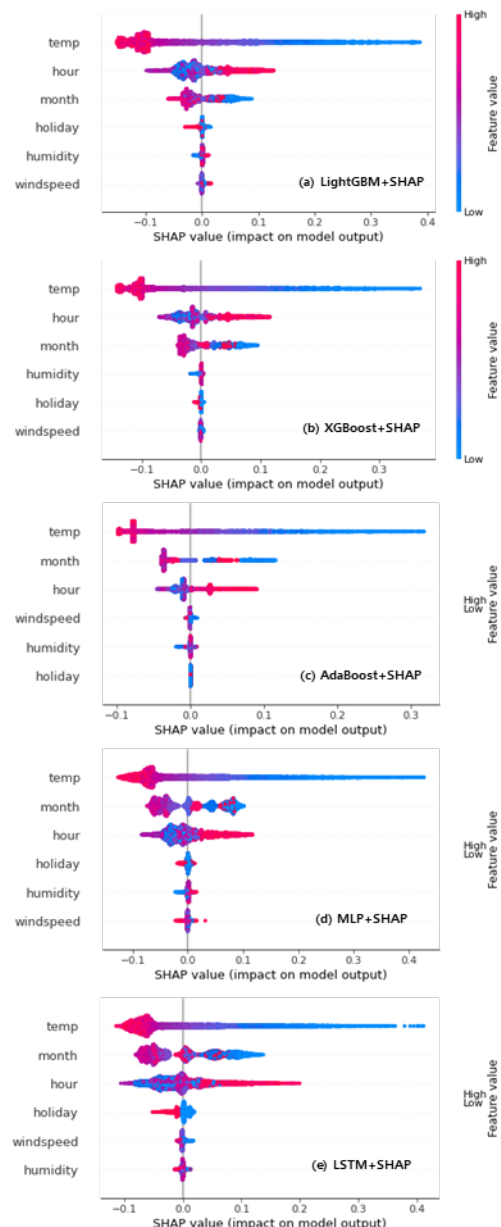
각 모델의 성능은 평가지표를 이용하여 측정한다. 평가지표는 R2 score와 RMSE, 그리고 MAPE이다. 평가지표 기준은 R2 score의 경우 1에 가까울수록 좋은 모델이고 0에 가까울수록 나쁜 모델이다. 0미만의 값이 나오는 경우 평균값으로 예측한 것보다 더 정확도가 떨어지는 예측을 하는 나쁜 모델이라고 할 수 있다. RMSE는 수치가 낮을수록 좋은 모델이라고 할 수 있다.

MAPE는 0~100% 사이의 값을 가지며 0에 가까울수록 좋은 모델이다. <표 2>는 각 모델의 성능을 평가한 결과이며, 값은 소수점 넷째 자리에서 반올림하여 셋째 자리까지 표기하였다.

<표 2> 모델 성능 평가

모델	R2 score	RMSE	MAPE
LightGBM	0.923	249.663	0.172
XGBoost	0.915	274.931	0.185
AdaBoost	0.898	328.716	0.336
MLP	0.902	315.413	0.218
LSTM	0.933	216.774	0.166

연구 결과 LSTM의 성능이 가장 좋았고, 그 다음으로 LightGBM이 높은 성능을 보였다. LightGBM과 LSTM의 훈련에 소요되는 시간을 비교해보았을 때, LightGBM은 LSTM보다 최소 600배 이상 빠른 속도를 보였다. 성능면에서는 LSTM이, 속도면에서는 LightGBM이 우수한 성능을 보였다.



<그림 4> 모델별 SHAP 적용 결과

XAI 는 eXplainable Artificial Intelligence 의 줄임말로 black-box 로 인해 볼 수 없는 모델을 설명하기 위해 등장한 개념이다[9]. 본 연구에서는 XAI 의 대표적인 모델 중 하나인 SHAP[10]을 이용한다. SHAP 은 모델에 사용된 특징(feature)이 결과 미치는 영향을 SHAP values 로 나타낸다. SHAP values 는 게임이론 개념을 이용하여 특징 중요도(Feature Importance)를 구하는 척도이다. <그림 4>는 각 모델별로 SHAP 을 적용하여 얻은 요약도(summary plot)이다. 모델별 요약도에서 y 축에는 모델에 미치는 특징의 중요도가 높은 순서대로 위에서 아래로 나타나며, x 축의 SHAP values 는 양수일 때 출력 변수가 높아지는데 기여를 하고, 음수일 때 출력 변수가 낮아지는데 기여를 한다. 각 모델의 특징 중요도 순위를 살펴보면 모두 기온(temp)이 가장 높은 기여를 하고 있는 특징임을 보여주며, 높은 기온일수록 열판매량(열공급량)이 낮아지는데 기여를 하고, 낮은 기온일수록 열판매량(열공급량)이 높아지는데 기여를 한다고 설명하고 있다. 이러한 방식을 통해 모델에 영향을 미치는 특징 중요도를 파악하고 변수를 선정하는 데 활용하고자 한다.

6. 결론

우리는 본 연구에서 열에너지 도메인으로 수요 예측을 위해 데이터셋을 구축하고 효율성 높은 모델 개발을 통해 방송과 미디어에 반드시 필요한 에너지 효율성 제고를 제안한다. 우리가 개발한 프레임워크는 단일 에너지 도메인에 국한하지 않으며 전력 소비의 효율성을 높이기 위한 설명적 모델 적용이 가능하다. 시간 단위로 10 년동안 수집된 난방 에너지 데이터와 수요 예측에 영향을 미치는 변수들을 선별하여 데이터셋을 구축하고 데이터 전처리를 위해 결측치 처리 및 정규화를 한다. 정확성을 보장하기 위해 LightGBM, XGBoost, AdaBoost, MLP, LSTM 5 가지 모델에 대한 성능 평가 비교를 진행하였고 LightGBM 과 LSTM 이 적합하다는 결과를 얻었다. 그리고 훈련된 모델을 설명 가능한 인공지능 모델 중 하나인 SHAP 을 도입하여 특징 중요도를 보고 모델에 영향을 미치는 특징과 관계에 대해서 파악했다. 이는 AI 모델 적용의 보편성의 근거가 되며 방송과 미디어에 인공지능 기술을 적극적으로 적용할 수 있는 플랫폼이 된다.

향후에는 XAI 모델을 이용하여 파악한 특징 중요도와 관계를 기반으로 하여 모델에 특징(feature)모델 성능 향상에 영향을 미치는 변수는 계속해서 사용하고, 모델 성능 하락에 영향을 미치는 변수를 제거하는 등의 방식을 통해 모델의 성능을 향상시키기 위한 변수 선정 방법에 대한 연구를 진행할 예정이다.

Acknowledgement

이 논문은 산업통산자원부 "열공급망 수용가 빅데이터 기반 에너지관리 기술개발 및 실증" 사업의 연구비지원 (20212020900150)에 의해 수행되었습니다.

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2022-0-00106, 복합 도메인의 에너지 수요 빅데이터를 활용한 설명 가능한 인공지능 기반 진단 및 분석 프레임워크 개발)

참고 문헌

- [1] Rajan, Ajitha, Adel Noureddine, and Panagiotis Stratis. "A study on the influence of software and hardware features on program energy." Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. 2016.
- [2] 송상화, et al. "실시간 열량계 정보를 활용한 단기 열 수요 예측 모델 개발에 관한 연구." 한국빅데이터학회지 5.2 (2020).
- [3] 최광용, and 권원태. "우리나라 사계절 개시일과 지속기간." 대한지리학회지 41.4 (2006).
- [4] Ke, Guolin, et al. "Lightgbm: A highly efficient gradient boosting decision tree." Advances in neural information processing systems 30 (2017).
- [5] 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. 2016.
- [6] Freund, Yoav, Robert Schapire, and Naoki Abe. "A short introduction to boosting." Journal-Japanese Society For Artificial Intelligence 14.771-780 (1999): 1612.
- [7] Minsky, Marvin, and Seymour Papert. "Perceptrons." (1969).
- [8] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.
- [9] Adadi, Amina, and Mohammed Berrada. "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)." IEEE access 6 (2018): 52138-52160.
- [10] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." Advances in neural information processing systems 30 (2017).