

주거용 건물의 에너지 소비량 예측을 위한 시퀀스 투 시퀀스 모델

Sequence-to-Sequence model for residential Building Energy consumption prediction

김윤재
세종대학교
컴퓨터공학과
서울시, 대한민국
jaiyun@sju.ac.kr

류상현
세종대학교
인공지능학과
서울시, 대한민국
22110777@sju.ac.kr

Nadeem Muhammad
세종대학교
컴퓨터공학과
서울시, 대한민국
22110104@sju.ac.kr

신지혜
세종대학교
인공지능학과
서울시, 대한민국
21110689@sju.ac.kr

권기학
세종대학교
컴퓨터공학과
서울시, 대한민국
kevinkkh@sejong.ac.kr

문현준*
세종대학교
컴퓨터공학과
서울시, 대한민국
hmoon@sejong.ac.kr

요약

최근 에너지 소비의 증가와 함께 효율적인 에너지 사용을 위해 에너지 소비량을 예측하는 기술들의 필요성이 증가하고 있다. 에너지 소비량을 예측하는 방법은 크게 2 가지로 분류할 수 있으며, 본 연구에서는 2 가지 방법 중 인공지능 기반의 접근 방식을 사용하여 DHW(Domestic Hot Water) 소비량을 예측하는 모델을 제안한다. DHW 소비량 예측 모델을 제안하기 위하여 XGBoost, LightGBM, LSTM 과 LSTM 을 활용한 Seq2Seq, 특징 추출을 위해 Seq2Seq 모델에 CNN 을 추가한 CNN-Seq2Seq, 총 5 가지 모델을 비교하였으며, 평가지표는 MAE, MSE, R^2 를 사용하였다. 입력 변수로는 기상청에서 획득한 기상 데이터와, 월, 일, 시와 같은 시간 데이터를 사용하였다. 평가지표 결과 CNN-Seq2Seq 모델이 (MAE)0.0810, (MSE)0.0204, (R^2)0.9701 로 5 가지 모델 중 제일 좋은 성능을 보여주었다.

키워드: DHW consumption, LSTM, Seq2Seq, Input Variable, CNN

1. 서론

도시화의 확산과 더불어 에너지의 소비는 지속적으로 증가하고 있으며, 전 세계 에너지 소비에서 건물이 차지하는 비중은 30% 이상이다 [1].

대한민국의 누적 온실가스 배출량 비율은 1960 년 이후 1.3%로 전 세계에서 16 위를 기록했으며, 2018 년 온실가스 배출량이 전 세계에서 11 번째로 많은 것으로 나타났다. 환경부에 따르면 대한민국 온실가스 배출량 중 에너지 분야가 1 위를 차지하고 있으며, 2 위인 산업공정 분야와 비교했을 때 약 12 배 높은 수치임을 알 수 있다. 또한 국토교통부에 따르면 에너지 분야 중 건물이 전체 에너지 소비량의 60%를 차지함을 알 수 있다.

건물에서의 에너지 소비가 증가함에 따라 에너지 소비 패턴에 대한 분석이 지속적으로 요구되고 있는 상태이며, 건물 에너지의 효율을 높일 수 있는 기술에 대한 관심이 높아지고 있다 [2].

건물의 에너지 소비량은 거주자가 얼마나 머무르는가, 거주자의 행동 패턴, 냉난방 시스템 등 다양한 요인에 의해 영향을 받는다 [3]. 따라서 건물에서 소비되는 에너지량을 예측할 수 있다면 에너지 공유를 위한 합리적인 소비로 에너지 저감에 영향을 줄 수 있다 [4].

에너지 소비량을 예측하는 방법은 크게 2 가지 범주로 분류할 수 있는데, 공학적인 접근 방법과 인공지능 기반의 접근 방식이다 [5].

공학적인 접근 방법은 수학적으로 측정 대상에 대한 모델링을 필요로 하는데, 이 과정이 매우 어려우며 많은 시간을 필요로 한다 [6]. 따라서 ANN (Artificial Neural Network)과 머신 러닝 모델들을 기반으로 한 인공지능 접근 방식이 많이 사용되고 있다 [7].

본 연구에서는 5 년 동안 수집된 DHW (Domestic Hot Water) 소비량 데이터를 XGBoost, LightGBM, LSTM, Seq2Seq, CNN-Seq2Seq, 총 5 가지의 모델에 적용하여 얻은 평가지표를 바탕으로 최적의 모델을 선정하였다.

2. 실험방법

2.1. 데이터셋

본 연구에서는 경기도 성남시에 위치한 16 동, 918 개의 세대가 존재하는 공동 주거용 아파트 시설에서 2017 년 11 월 1 일 00 시부터 2021 년 6 월 30 일 10 시까지 시간별로 수집된 총 29,163 개의 DHW 소비량 데이터를 사용하였다. 또한 기상청으로부터 기온, 지면온도, 풍속, 풍향,

습도, 시정, 강수, 해면기압, 일사, 일조, 운량을 제공받아 입력변수로 사용하였다.

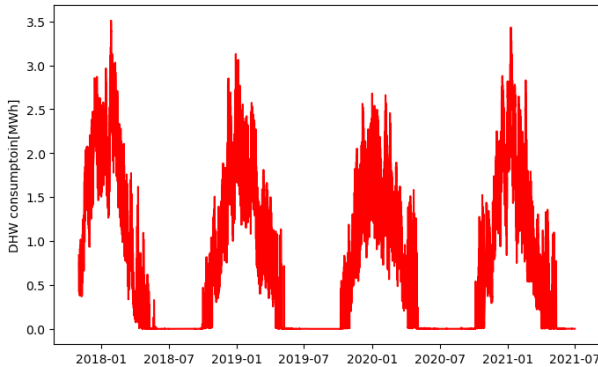


그림 1. DHW 소비량

수집된 DHW 소비량은 겨울에 해당하는 12월부터 2월까지 사용량이 증가하는 추세이며, 여름에 해당하는 6월부터 8월까지 급격히 감소하는 추세이다.

2.2. 실험환경

본 연구에서는 DHW 소비량 예측을 위해서 훈련 데이터셋과 테스트 데이터셋을 70%와 30%의 비율로 분리하였고, MinMaxScaler를 사용하여 정규화를 진행하였다.

Train dataset의 기간은 전체 데이터 기간의 70%인 2017년 11월 1일부터 2020년 6월 3일 오전 10시까지이고, Test dataset의 기간은 전체 데이터 기간의 30%에 해당하는 2020년 6월 3일 11시부터 2021년 6월 30일 오전 10시까지이다.

2.3. 모델

Gradient Boosting Decision Tree 기반의 분류 및 회귀 모델인 XGBoost는 우수한 예측 정확도와 효율적인 성능으로 건물 냉방 부하 예측 등 다양한 분야에 적용되는 알고리즘이다 [8-10]. LightGBM 역시 Gradient Boosting Decision Tree를 기반으로 한 앙상블 기계 학습 알고리즘으로 효율적인 작업과 높은 성능으로 인해 다양한 분야에서 광범위하게 사용되고 있다.

LSTM은 순환신경망의 단점을 보완한 모델로,

시계열 데이터를 분류, 처리 및 예측하는데 적합하다[9]. 본 연구에서는 DHW 소비량 데이터의 시계열 특성을 고려하여 XGBoost, LightGBM, LSTM을 채택하였으며 LSTM을 활용한 Seq2Seq, CNN-Seq2Seq 모델을 사용하였다. CNN-Seq2Seq 모델은 그림 2와 같다. 또한 LSTM은 3개의 layer를 쌓은 구조이며, Seq2Seq 모델은 CNN-Seq2Seq 모델 구조에서 입력 데이터의 특징 추출을 위한 CNN layer가 빠진 구조이다.

우리는 해당 모델들의 성능 비교를 위해 MAE, MSE, R^2 를 비교하였다.

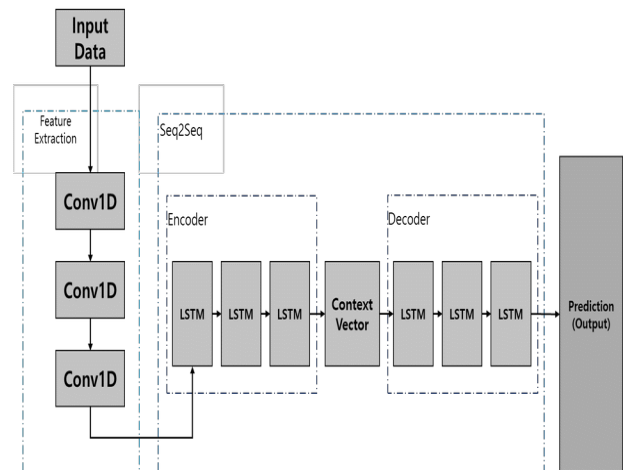


그림 2. CNN-Seq2Seq 모델 구조

Seq2Seq 모델은 Encoder와 Decoder로 구분할 수 있는데 Encoder는 input sequence를 통해 Context Vector를 만들고, Context Vector가 Decoder의 입력으로 들어가 과거 데이터를 기반으로 미래를 예측하는 구조이다.

본 논문에서는 Encoder에 3개의 LSTM layer와 Decoder에 3개의 LSTM layer를 연결하는 Seq2Seq 모델을 구축했으며, Input data에 CNN을 연결하여 특징 추출을 진행하였다.

3. 실험결과

본 실험 결과는 표 1과 같다.

표 1. 모델 별 평가지표 결과

	MAE	MSE	R^2
XGBoost	0.0961	0.0254	0.9624
LightGBM	0.0836	0.0210	0.9692

LSTM	0.0915	0.0281	0.9588
Seq2Seq	0.0900	0.0287	0.9579
CNN-Seq2Seq	0.0810	0.0204	0.9701

Seq2Seq 모델에 CNN 을 추가한 모델인 CNN-Seq2Seq 모델이 제일 좋은 성능을 보였다.

각 모델 별 예측 결과는 다음과 같으며 x 축은 Test dataset 에 해당하는 DHW 사용일자를 의미한다.

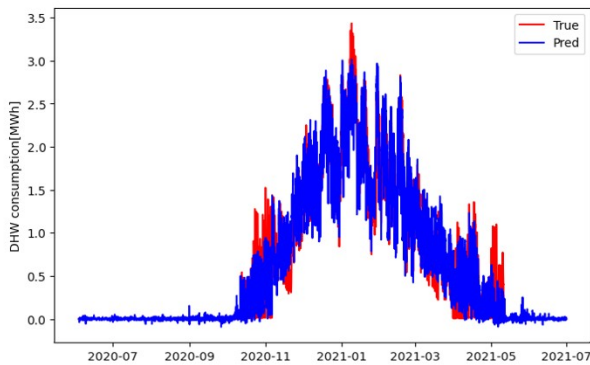


그림 3. XGBoost 예측 결과

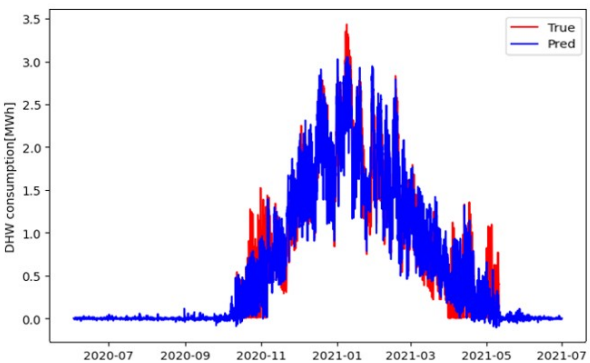


그림 4. LightGBM 예측 결과

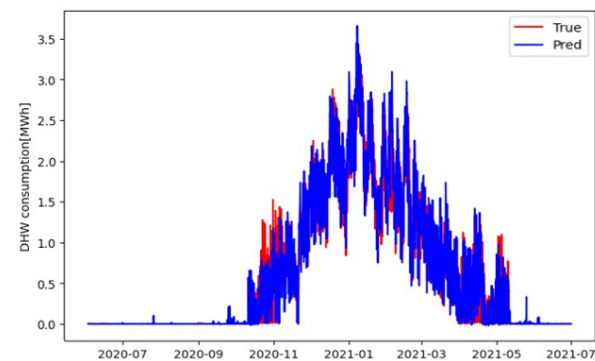


그림 5. LSTM 예측 결과

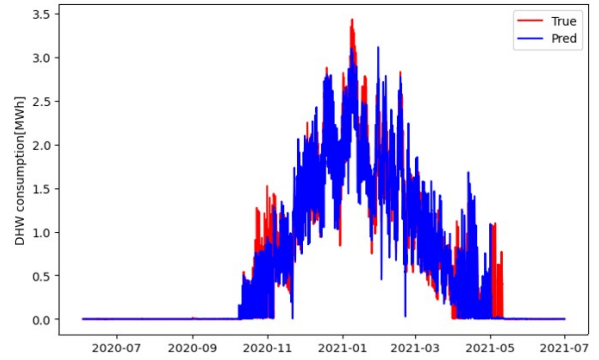


그림 6. Seq2Seq 예측 결과

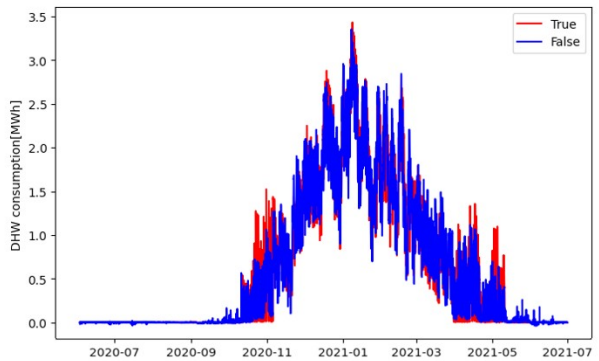


그림 7. CNN-Seq2Seq 예측 결과

4. 결론

본 논문에서는 DHW 소비량 예측을 위한 예측 모델 중 CNN-Seq2Seq 모델을 제안하였다. DHW 소비량 예측을 위해 수집된 DHW 소비량 데이터셋과 기상 데이터셋을 사용하였다. DHW 소비량 예측 모델의 성능을 비교하기 위해 XGBoost, LightGBM, LSTM, Seq2Seq, CNN-Seq2Seq 모델에 평가지표인 MAE, MSE, R^2 를 사용하여 비교하였다. 사용된 5 개의 모델 중 CNN-Seq2Seq 모델이 가장 낮은 MAE, MSE 와 제일 높은 R^2 인 0.9701 성능을 보여주었다. LSTM 과 Seq2Seq 모델의 예측 결과는 매우 비슷하지만, CNN 을 추가하여 Feature Extraction 을 진행한 CNN-Seq2Seq 모델이 제일 좋은 성능을 보여줌에 따라 CNN 을 이용한 Feature Extraction 이 예측 성능을 개선하는데 기여한 것으로 판단된다. 향후에는 해당 모델에 경량화 작업을 거친 후 XAI 를 적용하여, 어떤 변수들이 예측 시에 어떠한 영향을 주었는지에 대해 추가 연구를 진행할 예정이다.

Acknowledgement

이 논문은 정부(과학기술정보통신부)의 재원으로
한국연구재단의 지원을 받아 수행된 연구임(No.
2021R1F1A1046339).

이 논문은 산업통산자원부 “열 공급망 수용가
빅데이터 기반 에너지관리 기술개발 및 실증”
사업의 연구비지원(20212020900150)에 의해
수행됨.

이 논문은 2023 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)

참 고 문 헌

- [1] Ministry of Land, Infrastructure and Transport. (2014). A study on the standarization of quantitative evaluation method of building energy performance.
- [2] Al-Honmoud, M. S. (2001). Computer-aided building energy analysis techniques, Building and Environment.
- [3] A.M. Castro Martinez., S.H. Mallidi & B.T. Meyer. (2017). On the relevance of auditory-based Gabor features for deep learning in robust speech recognition, Computer Speech & Language, 45, 21-38.
- [4] Xiao, F. and Fan, C., 2014, Data Mining in Building Automation System for Improving Building Operational Performance, Energy Build, Vol. 75, No. 11, pp. 109-118.
- [5] Mariano-Hernández, D., Hernández-Callejo, L., García, F. S., Duque-Perez, O., & Zorita-Lamadrid, A. L. (2020). A review of energy consumption forecasting in smart buildings: methods, input variables, forecasting horizon and metrics. Applied Sciences, 10(23), 8323.
- [6] WANG, Zeyu; WANG, Yueren; SRINIVASAN, Ravi S. A novel ensemble learning approach to support building energy use prediction. Energy and Buildings, 2018, 159: 109-122.
- [7] Pavlicko, M.; Vojteková, M.; Blažeková, O. Forecasting of Electrical Energy Consumption in Slovakia. Mathematics 2022, 10, 577. <https://doi.org/10.3390/math10040577>.
- [8] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794.
- [9] J. Fan, X. Wang, L. Wu, H. Zhou, F. Zhang, X. Yu, X. Lu, Y. Xiang, Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global Solar radiation using Temperature and precipitation in humid subtropical climates: a case study in China, Energy Convers. Manag. 164 (2018) 102-111.
- [10] Y.-H. Chen, J.-L. Chen, Ai@ ntiphish—machine learning mechanisms for cyber-phishing attack, IEICE Trans. Info Syst. 102 (5) (2019) 878-887