LSTM을 사용한 SHAP 기반의 설명 가능한 태양광 발전량 예측 기법

박성우, 노윤아, 정승민, 황인준
고려대학교 전기전자공학과
psw5574@korea.ac.kr, yoona@korea.ac.kr, jmkstcom@korea.ac.kr, ehwang04@korea.ac.kr

SHAP-based Explainable Photovoltaic Power Forecasting Scheme Using LSTM

Sungwoo Park, Yoona Noh, Seungmin Jung, Eenjun Hwang
School of Electrical Engineering, Korea University

요 약

최근 화석연료의 급격한 사용에 따른 자원고갈이나 환경오염과 같은 문제들이 심각해짐에 따라 화석연료를 대체할 수 있는 신재생에너지에 대한 관심이 높아지고 있다. 태양광 에너지는 다른 에너지원에 비해 고갈의 우려가 없고, 부지 선정의 제약이 크지 않아 수요가 증가하고 있다. 태양광 발전 시스템에서 생산된 전력을 효과적으로 사용하기 위해서는 태양광 발전량에 대한 정확한 예측 모델이 필요하다. 이를 위한 다양한 딥러닝 기반의 예측 모델들이 제안되었지만, 이러한 모델들은 모델 내부에서 일어나는 의사결정 과정을 들여다보기가 어렵다. 의사결정에 대한 설명이 없다면 예측 모델의 결과를 완전히 신뢰하고 사용하는 데 제약이 따른다. 이런 문제를 위해서 최근 주목을 받는 설명 가능한 인공지능 기술을 사용한다면, 예측 모델의 결과 도출에 대한 해석을 제공할 수 있어 모델의 신뢰성을 확보할 수 있을 뿐만 아니라 모델의 성능 향상을 기대할 수도 있다. 이에 본 논문에서는 Long Short-Term Memory(LSTM)을 사용하여 모델을 구성하고, 모델에서 어떻게 예측값이 도출되었는지를 SHapley Additive exPlanation(SHAP)을 통하여 설명하는 태양광 발전량 예측 기법을 제안한다.

1. 서론

교토의정서보다 더욱 강력하게 온실가스 배출을 규제하는 조약인 파리협정의 효력이 2021년 1월 1일부터 발생하였다. 파리협정은 산업화 이전 수준 대비 지구 평균 온도가 1.5℃ 이상 상승하지 않도록 온실가스 배출량을 단계적으로 감축하는 것을 목표로 한다[1]. 온실가스 배출은 인간의 생산 및 소비 활동과 밀접하게 연관되어 있는데, 에너지, 산업공정, 농업, 폐기물 등의 다양한 배출원 중 에너지 부문과의 연관성이 가장 크다[2]. 에너지 부문에서 배출되는 온실가스의 양은 전체 배출량의 87%에 달한다.

온실가스 배출량을 줄이기 위해서는 에너지 부문에서 화석연료 사용량을 줄이고 다른 대체 에너지원의 사용량을 늘려야 한다. 하지만, 원자력의 경우 핵분열로 인하여 발생하는 방사능의 위험이 크고, 핵폐기물 처리에 많은 어려움이 따르고 있다. 이에, 많은 국가에서는 신재생 에너지 기술 개발에 많은 노력과 예산을 투입하고 있다[3]. 그중에서도 태양광 발전은 다른 에너지원에 비해 고갈의 염려가 없고, 부지 선정의 제약성이 크지 않는 등 다양한 장점으로 인하여 최근 관심과 수요가 급증하고 있다[4]. 이러한 태양광 발전 시스템에서 생산된 전력을 효과적으로 사용하기 위해서는 태양광 발전량에 대한 정확한 예측 모델이 필수적이다[5].

과거에는 이동 평균, 회귀 분석 방법 등과 같은 통계적 기법 기반의 예측 모델들이 많이 사용되었고, 최근에는 머신러닝 기법과 딥러닝 기반의 예측모델에 관한 연구가 활발히 진행되어, 좋은 예측 성능을 보여주고 있다[6]. 하지만 뛰어난 성능을 보여주는 딥러닝 기반의 모델의 경우, 모델 내부에서 일어나는 의사결정 과정을 확인할 수 없다는 문제가 존재한다. 의사결정 과정에 대한 설명이 없다면 예측모델을 통해 생성된 결과에 대하여 분석을 할 수 없으며, 그에 따라 모델을 완전히 신뢰하고 사용할 수 없게 된다. 이러한 문제점을 해결하기 위하여 설명 가능한 인공지능에 대한 연구가 활발하게 진행되고 있다[7].

설명 가능한 인공지능이란 사용자가 인공지능 모델의 학습 과정과 예측 결과를 해석하여 출력값이

생성되는 과정을 설명할 수 있도록 하는 기술을 말한다[8]. 인공지능 모델이 어떻게 결과를 도출하였는지를 해석할 수 있다면 모델의 신뢰성을 확보할 수 있을 뿐만 아니라 해석된 결과를 바탕으로 모델을 개선하여 성능을 향상할 수도 있다. 이에 본 논문에서는 Long Short-Term Memory(LSTM)을 사용하여 모델을 구성하고, 모델에서 어떻게 예측 값이 도출되었는지를 SHapley Additive exPlanation (SHAP)을 통하여 설명하는 태양광 발전량 예측 기법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 데이터 셋 구성에 대하여 설명하고, 3장에서는 제안하는 모델에 들어가는 기술들에 대하여 설명한다. 4장에서는 다양한 기계학습 기법 기반의 예측 모델과 제안한 기법의 성능을 비교·분석한다. 마지막으로 5장에서는 결론 및 향후 연구 방향으로 본 논문의 끝을 맺는다.

2. 데이터 셋

본 논문에서는 태양광 발전량 예측모델을 학습하기 위해 시간, 기상 데이터를 고려하였다. 태양광 발전량은 태양광 패널의 표면에 입사하는 일사량에 직접적인 영향을 받는다. 태양광 패널에 입사 가능한 일사량의 최대치는 시간 데이터의 영향을 받으며, 실제로 태양광 패널에 도달하는 일사량은 다양한 기상 데이터의 영향을 받는다.

시간 데이터는 월, 일, 시 데이터로 구성되어 있다. 시간 데이터를 1차원의 데이터 형태로 그대로 사용할 경우 시간의 주기성을 반영하기가 어렵다[9]. 따라서 주기함수를 통하여 2차원 데이터로 변환해준 후 2차원 데이터를 입력변수로 사용하였다.

기상 데이터는 기상청의 동네 예보에서 제공하는 일 평균 기온, 일 최저 기온, 일 최고 기온, 기온, 습도, 풍속, 강수량 데이터를 사용하였다. 표 1은 전체 입력 변수를 정리한 것이다.

<표 1> 예측 모델의 입력변수 목록

입력 변수	
Month_X	*Month_Y*
Day_X	*Day_Y*
Hour_X	*Hour_Y*
Average_Temperature	*Low_Temperature*
High_Temperature	*Temperature*
Humidity	*Wind_Speed*
Precipitation	

3. 모델 구성

3.1. LSTM

Recurrent Neural Network(RNN) 모델은 은닉층의 결과가 같은 은닉층의 입력으로 들어가는 순환적 구조로 되어 있어 시간적 순서를 가진 시계열 데이터 분석에 적합한 모델이다[10]. 하지만, RNN 모델은 시퀀스가 길어질수록 과거의 정보가 다음 단계로 잘 이루어지지 않아 학습 가중치가 사라져버리는 기울기 소실 문제가 발생한다. LSTM은 이러한 기울기 소실 문제를 해결하기 위하여 제안된 모델이다[11].

LSTM은 RNN과 동일하게 같은 네트워크를 반복하는 구조로 되어 있지만, Cell state를 추가로 가지고 있어 중요한 정보와 필요 없는 정보를 다음 state에 넘기며 장기기억을 보존한다. LSTM 모델은 Forget gate, Input gate, Output gate, 3개의 gate 구조를 거쳐서 Cell state에 정보를 저장한다. Forget gate에서는 어떤 정보를 버릴지 결정하고, Input gate에서는 어떤 값을 업데이트할 것인지를 결정하고, Output gate에서는 출력값을 결정한다.

3.2. SHAP

SHAP은 각각의 입력변수에 대한 SHAP value를 계산함으로써 입력 변수와 모델의 결과값 사이의 관계를 분석하는 설명 가능한 인공지능 기법이다[12]. SHAP value는 기계학습 모델의 조건부 기댓값 함수에 대한 섀플리 값으로 정의된다. 섀플리 값은 게임이론에서 게임의 참여자 간 협조로 얻어진 모든 이득으로부터 참여자들의 기여분에 따라 배분되는 값을 의미한다. 기계학습 모델에서는 예측 모델의 결과에 대한 변수 하나의 기여도를 모든 가능한 변수 조합에 대해서 종합적으로 구한 값을 의미한다. SHAP은 각각의 입력변수의 중요도를 확인할 수 있을 뿐만 아니라 결과에 미치는 영향력의 방향성과 크기까지 확인할 수 있어서 SHAP을 사용하여 모델을 분석하면 각각의 입력변수의 값이 변할 때 예측값이 어떻게 변화하는지를 해석할 수 있다.

4. 실험 결과

본 논문에서는 공공데이터포털에서 제공하는 인천, 부산 지역의 태양광 발전소에서 2018년 1월 1일부터 2019년 12월 31일까지 1시간 단위로 수집된 태양광 발전량 데이터를 사용하였다. 표 2는 각각의 장소에서 수집된 데이터의 통계적인 분석을 보여준다.

<표 2> 태양광 발전량 데이터 통계적 분석

	인천	부산
평균	244.88	225.17
표준편차	374.29	343.14
첨도	0.59	0.36
왜도	1.39	1.34
최소값	0	0
최대값	1479.6	1274.4
관측 데이터 수	17520	17520

4.1. 모델 정확도

제안한 모델의 성능을 평가하기 위해 다양한 기계학습 모델들과 비교하는 실험을 수행하였다. 실험에서는 2018년 1월 1일부터 2018년 12월 31일까지 수집된 데이터를 훈련용, 2019년 1월 1일부터 2019년 12월 31일까지 수집된 데이터를 테스트용으로 사용하였다.

기계학습 기법은 Linear Regression(LR), Decision Tree(DT), Support Vector Machine(SVM), Random Forest(RF), Gradient Boosting Machine(GBM), Adaptive Boosting(AdaBoost), MultiLayer Perceptron(MLP)으로 총 7가지를 사용하였으며, 예측 정확도로는 MAE(Mean Absolute Error)와 RMSE(Root Mean Square Error)를 사용하였다. 예측 성능을 비교한 결과는 표 3, 4와 같다.

<표 3> 각 데이터에 관한 MAE 비교

모델	인천	부산
LR	158.57	139.28
DT	96.71	71.33
SVM	101.56	88.02
RF	73.29	56.35
GBM	82.90	70.18
AdaBoost	98.53	107.49
MLP	90.25	78.30
Proposed Model	**67.33**	**53.45**

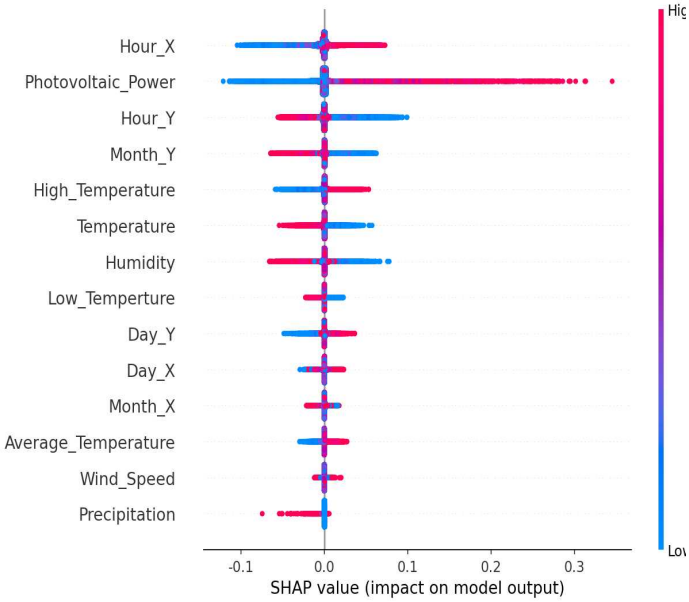
<표 4> 각 데이터에 관한 RMSE 비교

	인천	부산
LR	200.17	177.87
DT	203.48	162.55
SVM	151.92	122.81
RF	153.40	118.20
GBM	144.29	119.80
AdaBoost	170.60	166.21
MLP	146.76	122.78
Proposed Model	**127.79**	**99.72**

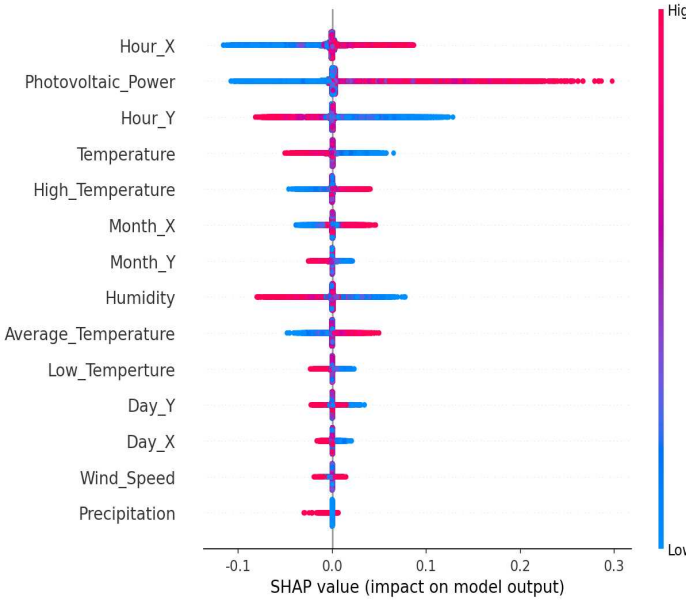
두 표에서 확인할 수 있듯이, 제안한 모델의 MAE나 RMSE 수치가 다른 기계학습 모델과 비교하였을 때 가장 좋은 결과를 보였다.

4.2. 모델 해석

그림 1, 2는 입력변수들의 SHAP value 값이 표시된 분포도이며, SHAP value의 절대값이 높은 순으로 나열되어 있다. 점의 색은 각 입력 변수의 값이 높고 낮은 정도를 나타내며, 중앙선을 기점으로 왼쪽의 점은 예측값이 낮아지는데, 오른쪽은 예측값이 높아지는 데 영향을 준다는 것을 의미한다.



(그림 1) 입력변수의 SHAP value 분포도(인천)



(그림 2) 입력변수의 SHAP value 분포도(부산)

그림 1과 2 모두에서 Hour_X의 영향력이 가장 중요하다는 결과를 얻을 수 있었으며 시 데이터와 기온 데이터가 태양광 발전량 예측에 있어서 중요하다는 것을 확인할 수 있었다. 또한, LSTM 모델의 입력으로 들어가는 과거 태양광 발전량 데이터의 영향력이 크다는 것을 확인할 수 있었다.

5. 결론

본 논문에서는 LSTM 모델을 구성하여 태양광 발전량을 예측하고 SHAP을 통하여 설명하는 설명 가능한 태양광 발전량 예측 기법을 제안하였다. 데이터 셋으로는 인천과 부산에 위치한 태양광 발전소의 1시간 단위 태양광 발전량 데이터를 사용하였다. 다양한 기계학습 기법과 예측 성능을 비교한 결과, 제안한 LSTM 기반의 예측 모델의 성능이 가장 뛰어남을 확인할 수 있었다.

또한, 인공지능 해석기법 중 하나인 SHAP를 통해 태양광 발전량 예측을 위한 각 입력변수에 대하여 분석하였다. 태양광 발전량 예측을 위한 각 입력변수에 대하여 SHAP Value 등을 계산하고 SHAP Value 들로부터 입력 특징 중요도를 산출하였다. 이를 분석한 결과 시 데이터와 기온 데이터가 태양광 발전량 예측에 있어서 중요한 역할을 한다는 것을 확인할 수 있었다.

향후, 본 연구에서 제안한 기법을 보다 국내 데이터뿐만 호주, 독일 등의 다양한 국가에서 수집된 태양광 발전량 데이터에 적용하여 예측을 진행할 계획이다. 해외의 기상조건에서는 어떤 변수가 가장 큰 영향을 주는지 확인함과 동시에 기상 조건별로 데이터를 군집화 한 뒤 군집별 예측 모델을 구성하고 데이터 영향력을 분석하여 이를 바탕으로 더욱 정확한 태양광 발전량 예측 모델을 구성할 계획이다.

사사문구

본 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단—에너지클라우드기술개발사업의 지원(No. 2019M3F2A1073184)을 받아 수행된 연구임.

참고문헌

[1] S. N. Seo, "Beyond the Paris Agreement: Climate change policy negotiations and future directions," Regional Science Policy & Practice, Vol. 9, No. 2, pp. 121-140, 2017.

[2] A. Karmaker er al., "Exploration and corrective measures of greenhouse gas emission from fossil fuel power stations for Bangladesh," Journal of Cleaner Production, Vol. 244, 2020.

[3] K. Hansenet al., "Full energy system transition towards 100% renewable energy in Germany in 2050," Renewable and Sustainable Energy Reviews, Vol. 102, pp. 1-13, 2019.

[4] M. Lee et al., "24-Hour photovoltaic generation forecasting using combined very-short-term and short-term multivariate time series model," In 2017 IEEE Power & Energy Society General Meeting, pp. 1-5, 2017.

[5] B. Zhang et al., "A multiple time series-based recurrent neural network for short-term load forecasting," Soft Computing, Vol. 22, pp. 4099-4112, 2017.

[6] J. Jeong et al., "Improvement for Forecasting of Photovoltaic Power Output using Real Time Weather Data based on Machine Learning," Journal of Korean Society of Living Environmental System, Vol. 25, pp. 119-125, 2018.

[7] D. Gunning, "Explainable artificial intelligence (XAI)," Defense Advanced Research Projects Agency (DARPA), 2017.

[8] D. H. Park et al., "Multimodal explanations: Justifying decisions and pointing to the evidence," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8779-8788, 2018.

[9] S. Park et al., "A two-stage industrial load forecasting scheme for day-ahead combined cooling, heating and power scheduling," Energies, Vol. 13, No. 2, pp. 443-465, 2020.

[10] Y. LeCun et al., "Deep learning," nature, Vol. 521, No. 7553, pp. 436-444, 2015.

[11] H. Sak et al., "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," 2014.

[12] S. M. Lundberg et al., "A Unified Approach to Interpreting Model Predictions," Advances in neural information processing systems, pp. 4765-4774, 2017.