

Explainable Photovoltaic Power Forecasting Scheme Using BiLSTM

Sungwoo Park[†] · Seungmin Jung^{††} · Jaek Moon^{††} · Eenjun Hwang^{†††}

ABSTRACT

Recently, the resource depletion and climate change problem caused by the massive usage of fossil fuels for electric power generation has become a critical issue worldwide. According to this issue, interest in renewable energy resources that can replace fossil fuels is increasing. Especially, photovoltaic power has gaining much attention because there is no risk of resource exhaustion compared to other energy resources and there are low restrictions on installation of photovoltaic system. In order to use the power generated by the photovoltaic system efficiently, a more accurate photovoltaic power forecasting model is required. So far, even though many machine learning and deep learning-based photovoltaic power forecasting models have been proposed, they showed limited success in terms of interpretability. Deep learning-based forecasting models have the disadvantage of being difficult to explain how the forecasting results are derived. To solve this problem, many studies are being conducted on explainable artificial intelligence technique. The reliability of the model can be secured if it is possible to interpret how the model derives the results. Also, the model can be improved to increase the forecasting accuracy based on the analysis results. Therefore, in this paper, we propose an explainable photovoltaic power forecasting scheme based on BiLSTM (Bidirectional Long Short-Term Memory) and SHAP (SHapley Additive exPlanations).

Keywords : Smart Grid, Photovoltaic Power Forecasting, Deep Learning, Explainable Artificial Intelligence

BiLSTM 기반의 설명 가능한 태양광 발전량 예측 기법

박성우[†] · 정승민^{††} · 문재욱^{††} · 황인준^{†††}

요약

최근 화석연료의 무분별한 사용으로 인한 자원고갈 문제 및 기후변화 문제 등이 심각해짐에 따라 화석연료를 대체할 수 있는 신재생에너지에 대한 관심이 증가하고 있다. 특히 신재생에너지 중 태양광 에너지는 다른 신재생에너지원에 비해 고갈될 염려가 적고, 공간적인 제약이 크지 않아 전국적으로 수요가 증가하고 있다. 태양광 발전 시스템에서 생산된 전력을 효율적으로 사용하기 위해서는 보다 정확한 태양광 발전량 예측 모델이 필요하다. 이를 위하여 다양한 기계학습 및 심층학습 기반의 태양광 발전량 예측 모델이 제안되었지만, 심층학습 기반의 예측 모델은 모델 내부에서 일어나는 의사결정 과정을 해석하기가 어렵다는 단점을 보유하고 있다. 이러한 문제를 해결하기 위하여 설명 가능한 인공지능 기술이 많은 주목을 받고 있다. 설명 가능한 인공지능 기술을 통하여 예측 모델의 결과 도출 과정을 해석할 수 있다면 모델의 신뢰성을 확보할 수 있을 뿐만 아니라 해석된 도출 결과를 바탕으로 모델을 개선하여 성능 향상을 기대할 수도 있다. 이에 본 논문에서는 BiLSTM(Bidirectional Long Short-Term Memory)을 사용하여 모델을 구성하고, 모델에서 어떻게 예측값이 도출되었는지를 SHAP(SHapley Additive exPlanations)을 통하여 설명하는 설명 가능한 태양광 발전량 예측 기법을 제안한다.

키워드 : 스마트 그리드, 태양광 발전량 예측, 심층학습, 설명가능한 인공지능

1. 서론

현재 인류가 의존하고 있는 에너지 대부분은 화력 발전

으로 생성되고 있다[1]. 화력 발전은 석유나 석탄과 같은 화석연료를 연소하여 얻은 증기로 터빈을 돌려 전력을 생산하는 단순한 형태의 전력 생산 방식이다. 화력 발전은 0.05~0.177달러/kWh의 낮은 단가로 전력을 생산할 수 있지만[2], 화석연료의 연소 과정에서 다량의 이산화탄소나 질소 산화물, 미세먼지 등과 같은 유해 성분이 배출된다는 단점이 존재한다[3]. 또한, 화석연료는 매장량이 한정되어 있어서 언젠가는 고갈되어 터는 사용할 수 없다[4].

화석연료의 연소 과정에서 배출되는 유해 성분 중 이산화탄소는 대표적인 온실가스로 지구온난화의 주범으로 꼽히고 있다[5]. 온실가스는 지구의 지표면에서 우주로 발산하는 적

※ 본 연구는 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-에너지클라우드기술개발사업(No. 2019M3F2A1073184)의 지원을 받아 수행된 연구임.

※ 이 논문은 2021년 한국정보처리학회 ACK 2021에서 "LSTM을 사용한 SHAP 기반의 설명 가능한 태양광 발전량 예측 기법"의 제목으로 발표된 논문을 확장한 것임.

† 준회원 : 고려대학교 전기전자공학부 박사과정

†† 비회원 : 고려대학교 전기전자공학부 박사과정

††† 종신회원 : 고려대학교 전기전자공학부 교수

Manuscript Received : December 30, 2021

Accepted : January 17, 2022

* Corresponding Author : Eenjun Hwang(ehwang04@korea.ac.kr)

외선 복사열을 흡수 또는 반사하여 지구 표면의 온도를 생물들이 서식하는데 적합한 영상 15도로 유지하는 데 도움을 주는 기체이다. 하지만 온실가스가 필요 이상으로 증가하게 되면, 지구의 평균 기온이 상승하게 되는 지구온난화 현상이 발생하게 된다. 지구온난화로 인하여 해수면 상승 및 생태계 교란 등의 기후변화 문제가 발생할 수 있다[6].

이러한 기후변화 문제나 자원고갈 문제 등을 해결할 수 있는 수단으로 스마트 그리드 기술이 많은 주목을 받고 있다[7]. 스마트 그리드는 ICT(Information and Communication Technology)를 기존의 전력망에 접목한 차세대 지능형 전력망 구조를 의미한다. 스마트 그리드는 전력 공급 업체와 소비자 간에 실시간 정보를 교환함으로써 에너지 효율을 향상할 수 있다[8]. 또한, 스마트 그리드는 수소, 풍력, 바이오, 태양광 에너지와 같은 신재생 에너지에 바탕을 둔 분산 전원의 활성화를 통해 기존의 발전설비에 들어가는 화석연료의 사용량을 줄일 수 있다[9].

신재생 에너지를 이용한 발전 중 태양광 발전은 다른 에너지원에 비해 고갈될 염려가 전혀 없고, 부지 선정의 제약성이 크지 않으며, 초기 투자비용이 상대적으로 적다는 등의 장점으로 인하여 최근 관심 및 수요가 급증하고 있다[10]. 태양광 발전 시스템에서 생산된 전력을 효율적으로 사용하기 위해서는 정확한 태양광 발전량 예측 모델이 필수적이다. 정확한 예측 모델은 안정적인 전력 공급 및 운영을 가능하게 하며, 불필요한 에너지 손실을 줄임으로 인하여 경제적인 이득을 극대화 할 수 있다[11].

과거에는 태양광 발전량 예측 방법으로 통계적 기법들이 많이 사용되었고, 최근에는 기계학습 기법과 심층학습 기반의 예측 모델에 관한 연구가 활발히 진행되고 있으며 좋은 예측 성능을 보여주고 있다[12]. 하지만 뛰어난 성능을 보여주는 심층학습 기반의 예측 모델은 모델 내부에서 일어나는 의사결정 과정을 설명하기가 어렵다는 단점이 존재한다. 의사결정 과정에 대한 설명이 없다면 예측 모델을 통해 생성된 결과가 어떤 과정을 통해 도출 되었는지 알 수 없고, 그에 따라 모델의 신뢰성이 저하된다. 이러한 문제점을 해결하기 위하여 설명 가능한 인공지능에 대한 연구가 활발하게 진행되고 있다[13].

설명 가능한 인공지능이란 인공지능 모델의 학습 과정과 예측 결과를 해석하여 출력값이 생성되는 과정을 설명할 수 있도록 하는 기술을 말한다[14]. 인공지능 모델의 예측 결과 도출 과정을 해석할 수 있다면 모델의 신뢰성을 확보할 수 있을 뿐만 아니라 해석된 도출 과정 결과를 바탕으로 모델을 개선하여 성능 향상을 기대할 수도 있다. 이에 본 논문에서는 BiLSTM(Bidirectional Long Short-Term Memory)을 사용하여 모델을 구성하고, 모델의 예측값 도출 과정을 SHAP(SHapley Additive exPlanations)을 통하여 설명하는 설명 가능한 태양광 발전량 예측 기법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 태양광 발전량 예측과 관련된 연구를 소개하며, 3장에서는 제안하는 기법에 대하여 자세하게 설명한다. 4장에서는 다양한 기계학습 및

심층학습 기반의 예측 모델들과 제안한 기법의 성능을 비교 분석한다. 마지막으로 5장에서는 결론 및 향후 연구 방향으로 본 논문의 끝을 맺는다.

2. 관련 연구

2.1 통계적 기법 기반의 태양광 발전량 예측

과거 태양광 발전량 데이터의 통계적 특성을 통해 통계적 모델을 설계하는 것은 태양광 발전량 예측의 대표적인 접근법이다. Massida et al.[15]은 날씨 예보 데이터를 MARS(Multilinear Adaptive Regression Splines)의 입력으로 사용하여 독일 보쿰에 위치한 태양광 발전소의 하루 후의 발전량을 예측하였다. 적은 수의 샘플을 사용했음에도 불구하고 제안한 방법은 우수한 결과를 보여주었다. Halabi et al.[16]은 일조 지속 시간 및 기온과 같은 다양한 기상 변수를 활용한 퍼지 알고리즘 기반의 월별 태양광 발전량 예측 기법을 제안하였다. 제안하는 기법에는 PSO(Particle Swarm Optimization), GA(Genetic Algorithm), DE(Differential Evolution)가 포함되며, 다양한 평가지표에서 우수한 성능을 보여주었다.

2.2 기계학습 기반의 태양광 발전량 예측

최근 컴퓨팅 기술의 비약적인 발전으로 인하여 기계학습 및 심층학습 기반의 예측 모델에 대한 연구가 활발하게 이루어지고 있으며 통계적 기법보다 우수한 성능을 보여주고 있다. Jeong et al.[12]은 기상 데이터를 다양한 기계학습 모델들의 입력으로 사용하여 일사량 및 태양광 발전량을 예측하였다. 비교한 기계학습 모델에는 SVM(Support Vector Machine), RF(Random Forest) 등이 있고, RF가 가장 좋은 성능을 보였다. Frank et al.[17]은 IoT(Internet of Things) 시스템에 사용되는 태양광 발전량을 예측하기 위해 기상 데이터를 사용하였고, 다양한 기계학습 모델들의 성능을 비교하였다. DT(Decision Tree), KNN(K-Nearest Neighbor), SVM, RT(Random Tree), RC(Random Committee) 등의 성능을 비교한 결과, RT와 RC가 좋은 성능을 보여주었다.

2.3 심층학습 기반의 태양광 발전량 예측

Almonacid et al.[18]은 태양광 발전량 예측을 위해 ANN(Artificial Neural Network)을 사용하였고, 전통적인 수학적 모델들과의 비교를 통해 해당 모델의 우수성을 증명하였다. ANN 모델은 낮은 조도, 각도 및 스펙트럼 효과와 같은 일부 2차적인 효과까지 고려하기 때문이다. Wang et al.[19]은 풍속, 기온, 습도 등의 날씨 데이터를 수집한 후 LSTM(Long Short-Term Memory)을 통해 시간적 패턴을 임베딩하고, CNN(Convolutional Neural Network)을 통해 변수 사이의 관계를 학습하여 미래의 태양광 발전량을 예측하는 모델을 구성하였다. 제안한 심층학습 기반의 하이브리드 모델이 다양한 단일 모델보다 우수함을 보였다.

3. 데이터 수집 및 전처리

본 논문에서는 제주, 부산, 인천, 하동 지역의 태양광 발전소에서 수집된 태양광 발전량 데이터를 사용하였다. 데이터는 공공데이터포털에서 제공되는 공공 데이터이며, 2018년 1월 1일부터 2019년 12월 31일까지 1시간 단위로 수집되어 있다. Table 1은 각각의 장소에서 수집된 데이터의 통계적인 분석을 보여준다.

수집한 데이터를 통해 예측 모델을 구성하기 위해서는 태양광 발전량과 밀접한 관계를 갖는 다양한 외부 변수들을 고려해야 한다[20]. 본 논문에서는 보다 정확한 태양광 발전량 예측 모델을 구성하기 위해서 시간 데이터와 기상 데이터를 고려하였다. 태양광 발전량은 태양광 패널의 표면에 입사하는 일사량에 직접적인 영향을 받는데 태양광 패널에 입사 가능한 일사량의 최대치는 계절 및 시간에 따라 달라지고, 실제로 도달하는 일사량은 다양한 기상요소에 따라 달라지기 때문이다[21].

3.1 시간 데이터

시간 데이터는 월, 일, 시 데이터로 구성되어 있다. 시간 데이터는 1차원 데이터 형태로 이루어져 있는데, 이를 그대로 사용할 경우 시간 데이터의 주기성을 정확하게 반영하기가 어렵다. 예를 들어, 11월 30일 23시와 12월 1일 0시는 실제로는 인접한 날짜와 시간이지만, 일 데이터 1차원 형태 그대로 사용할 경우 29의 차이가, 시 데이터의 경우에는 23의 차이가 발생한다. 본 논문에서는 Equation (1)~(6)을 사용하여 1차원 형태의 월, 일, 시 데이터를 2차원 공간의 연속적인 데이터로 증강하여 주기성을 반영하였다.

$$Month_x = \sin(Month \times (2\pi/12)) \quad (1)$$

$$Month_y = \cos(Month \times (2\pi/12)) \quad (2)$$

$$Day_x = \sin(Day \times (2\pi/DoM_{Month})) \quad (3)$$

$$Day_y = \cos(Day \times (2\pi/DoM_{Month})) \quad (4)$$

$$Hour_x = \sin(Hour \times (2\pi/24)) \quad (5)$$

$$Hour_y = \cos(Hour \times (2\pi/24)) \quad (6)$$

Equation (3), (4)의 DoM_{Month} 는 Day of the Month로 속한 달의 일 수를 나타낸다. Month가 1인 경우는 DoM_{Month} 가 31, Month가 2인 경우는 28 또는 29, Month가 4인 경우는 30을 의미한다.

각 차원 데이터는 서로 다른 장점을 보유하고 있다. 1차원 데이터는 범주형 정보를 반영하는데 효과적이고, 2차원 데이터는 주기성 정보를 반영하는데 효과적이다. 본 논문에서는 각 장점을 모두 활용하기 위하여 1차원 데이터와 2차원 데이터를 모두 입력변수로 사용하였다[22]. 결과적으로 본 논문에서는 총 9개의 시간 데이터를 입력변수로 사용하였다.

3.2 기상 데이터

기상 데이터는 기상자료개방포털에서 제공하는 데이터를 사용하였다. 기상청은 초단기 예보, 단기 예보, 중기 예보, 장기 예보와 같은 다양한 형태의 기상 예보를 제공하고 있다. 초단기 예보는 기온, 강수, 낙뢰 등 기상요소의 현재 실태부터 4시간 이내의 예보를 1시간 간격으로, 단기 예보는 기온, 강수, 하늘 상태 등 12개의 기상 요소의 3일 이내 예보를 3시간 간격으로 발표한다.

태양광 발전량 예측 모델을 통해 적어도 하루 전에는 예측이 수행되어야 전력 사용 전략을 수립할 수 있기 때문에 기상청의 단기 예보에서 하루 뒤의 예상 수치를 제공하는 일 평균 기온, 일 최고 기온, 일 최저 기온, 기온, 습도, 풍속, 강수량을 사용하였다. Fig. 1은 기상청의 단기 예보에서 제공하는 예보현황을 보여준다.

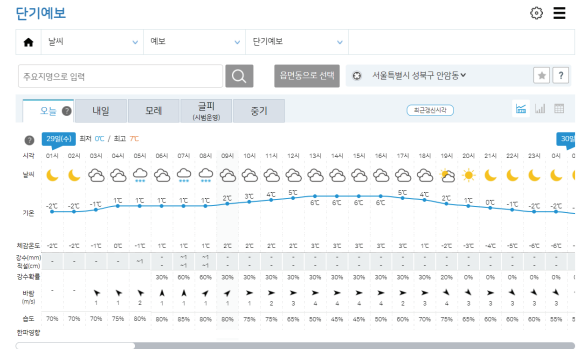


Fig. 1. Example of Short-Term Weather Forecast Information

Table 1. Statistical Analysis of Photovoltaic Power Data in the Four Locations

	Location			
	Jeju	Incheon	Busan	Hadong
Mean	23.5836	244.8796	225.1691	5.9807
Standard Error	38.8113	374.2908	343.1427	9.4336
Kurtosis	1.3608	0.5945	0.3633	0.7225
Skewness	1.6097	1.3915	1.3424	1.4505
Minimum	0	0	0	0
Maximum	157.359	1479.6	1274.4	38.004
Count	17520	17520	17520	17520

태양광 발전량 데이터는 1시간 단위로 수집된 반면 기상청의 단기 예보 데이터는 3시간 단위로 제공된다. 본 논문에서는 단기 예보에서 제공하는 데이터를 선형보간법을 사용하여 1시간 단위의 데이터로 변환시켜준 후 사용하였다.

4. 태양광 발전량 예측 모델 구성

이 장에서는 태양광 발전량 예측 모델을 이루고 있는 기술들에 대하여 보다 자세히 설명한다. 먼저 앞서 구성한 16개의 입력변수 중 2차원 형태의 월, 일, 시 데이터를 제외한 10개의 데이터들을 Equation (7)과 같이 최소-최대 정규화 방법을 통해 0과 1 사이의 값으로 정규화를 시켜준다.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (7)$$

정규화를 수행해주면 심층학습 모델의 초기화가 어떤 값으로 이루어지더라도 쉽게 최적값에 도달할 수 있으며, 학습률을 상대적으로 높여서 사용할 수 있기 때문에 빠르게 훈련이 가능하다[23].

4.1 BiLSTM

RNN(Recurrent Neural Network) 모델은 은닉층의 결과가 다음 단계의 은닉층의 입력으로 들어가는 순환적 구조로, 시계열 데이터 분석에 적합한 모델이다[24]. 하지만, RNN 모델은 시퀀스가 길어질수록 과거의 정보가 다음 단계로 잘 전달되지 않아 단계가 거듭될수록 학습 가중치가 사라져버리는 기울기 소실 문제가 발생할 수 있다.

LSTM은 이러한 RNN의 기울기 소실 문제를 해결하기 위하여 제안된 모델이다. LSTM은 Cell state를 통해 중요한 정보와 필요 없는 정보를 구분하여 다음 state에 넘기며 단계가 거듭되어도 장기기억을 보존할 수 있다[25]. LSTM 모델은 3개의 gate 구조를 거쳐서 Cell state에 정보를 저장한다. Input gate에서는 어떤 값을 업데이트할 것인지를 결정하고,

Output gate에서는 출력 값을 결정하며, Forget gate에서는 어떤 정보를 버릴지를 결정한다. Fig. 2는 LSTM의 구조를 나타낸다.

LSTM은 뛰어난 성능을 제공함에도 불구하고 여전히 입력 순서를 시간 순서대로 처리하기 때문에 결과는 주로 직전 패턴을 기반으로 수립하는 경향을 보인다는 한계가 있다. 이러한 단점을 해결하기 위해서 BiLSTM(Bidirectional Long Short-Term Memory)이 제안되었다. BiLSTM은 입력과 출력 계층에 순방향 hidden layer와 역방향 hidden layer이 연결된 구조를 보이며, 양방향으로 학습을 진행하고 각각의 hidden layer에서 나온 출력값을 연산하여 최종적인 출력값을 도출한다[26]. 순방향과 역방향으로 각각 학습이 진행된 후 합쳐지기 때문에 기존의 단방향 LSTM에 비해 더 좋은 예측 성능을 기대할 수 있다. Fig. 3은 BiLSTM의 구조를 나타낸다.

4.2 SHAP

복잡한 학습모델의 경우 각각의 입력변수가 얼마나 많은 영향을 끼쳐서 결과가 도출되었는지 해석하기가 어렵다. 따라서 원래 모델의 예측값을 해석 가능한 모델로 근사하는 방법이 필요하다. SHAP은 모델의 예측값과 입력변수 간의 관계를 탐색하기 위한 새로운 접근 방식이다[27]. SHAP은 각 입력변수에 대한 SHAP Value를 계산하여 입력변수를 분석하는 방법을 제공한다. SHAP Value는 기계학습 모델의 조진부 기댓값 함수에 대한 샤프리 값으로 정의된다.

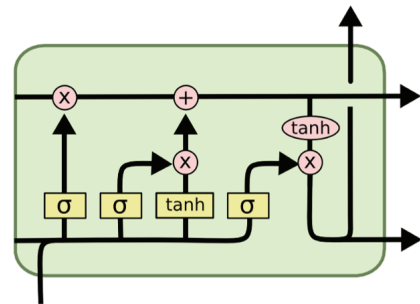


Fig. 2. LSTM Structure

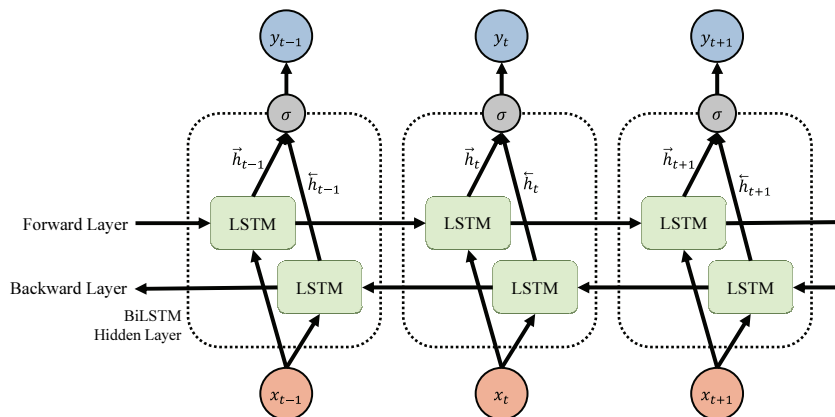


Fig. 3. BiLSTM Structure

기존에 기계학습 앙상블 모델에서 활용되는 변수 중요도를 통해 각 입력 변수가 결과에 미치는 영향력을 알 수 있었지만, 영향의 방향성을 알 수 없고, 입력 변수의 관계가 독립적이지 않고 의존성이 조금이라도 존재할 경우 그 결과를 신뢰할 수 없다는 단점이 있었다. SHAP은 이러한 단점들을 모두 극복하여 보다 논리적인 방식으로 입력 변수들이 결과에 미치는 영향을 계산할 수 있다. 또한, 도출된 SHAP value는 여러 번 반복 수행하여 얻어진 결과물의 평균이기에 일관성이 있어 더욱 신뢰하고 사용할 수 있다. 따라서, SHAP을 사용하여 모델을 분석하면 각각의 입력변수의 값이 변할 때 예측값이 어떻게 변화하는지를 해석할 수 있다.

5. 실험 결과

5.1 모델 정확도

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

실험에 사용한 기계학습 기법 및 심층학습 기법은 LR

(Linear Regression), DT, SVM, RF, GBM(Gradient Boosting Machine), AdaBoost(Adaptive Boosting), ANN, RNN, LSTM으로 총 9가지를 사용하였으며, 예측 정확도로는 MAE (Mean Absolute Error)와 RMSE(Root Mean Square Error)를 사용하였다. MAE는 모델의 오차를 절댓값의 평균으로 계산하여 실제값과의 차이에 대해 직관적인 해석이 가능하다는 장점이 있고, RMSE는 오차에 제곱근을 씌워 큰 오차가 발생했을 때 큰 패널티를 주어서 이상치에 큰 영향을 받지 않는 장점이 있다. MAE와 RMSE를 수식으로 표현하면 Equation (8), (9)와 같다.

$$MAE = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (9)$$

Equation (8), (9)에서 Y_i 는 실제 태양광 발전량 데이터를 의미하고 \hat{Y}_i 는 예측값을 의미한다. 예측 성능을 비교한 결과는 Table 2, Table 3과 같다.

Table 2. MAE Comparison for Photovoltaic Power Data in the Four Locations

Model	Location			
	Jeju	Incheon	Busan	Hadong
LR	18.7388	157.1073	139.1566	3.7478
DT	11.7624	97.7581	71.6067	1.9702
SVM	13.2790	104.7527	89.1074	2.5077
RF	9.4423	72.4535	55.6445	1.5741
GBM	10.7896	83.2068	69.3370	2.0116
ADA	12.4238	97.5441	92.4361	3.0800
ANN	13.0040	86.9994	76.3991	2.5072
RNN	7.9003	69.5959	59.1321	1.5645
LSTM	7.7757	63.1466	52.8873	1.5006
BiLSTM(proposed)	7.1187	61.8656	51.2317	1.4785

Table 3. RMSE Comparison for Photovoltaic Power Data in the Four Locations

Model	Location			
	Jeju	Incheon	Busan	Hadong
LR	23.8253	199.3280	178.1562	4.8586
DT	24.7759	206.6747	164.1063	4.3901
SVM	20.7146	156.9751	126.7972	3.6298
RF	19.2114	151.8863	119.1258	3.4020
GBM	18.9519	145.4381	119.7673	3.4883
ADA	21.1549	169.4445	154.5621	4.3813
ANN	21.0755	145.8452	121.4644	3.8001
RNN	13.6955	116.0512	99.2613	2.7743
LSTM	13.5978	113.6877	95.8080	2.7622
BiLSTM(proposed)	13.2536	113.0661	95.3631	2.7525

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

5.2 모델 해석

제안하는 BiLSTM 기반의 모델을 SHAP의 summary plot을 통해 해석한 결과를 Fig. 4에 나타내었다. Summary

plot은 입력변수들의 SHAP value 값을 절대값이 높은 순으로 나열한 분포도이다. 파란색에서 빨간색으로 갈수록 입력 변수의 값이 커진다는 것을 의미하고, 중앙선의 왼쪽에 분포한 점들은 최종 예측값을 작게, 오른쪽 점들은 예측값을 크게 도출하는데 영향을 주는 것을 의미한다.

Fig. 4를 해석한 결과, (a), (b), (c), (d) 모두에서 시간 데이터에 해당하는 Hour_X가 예측 결과를 도출하는데 가장

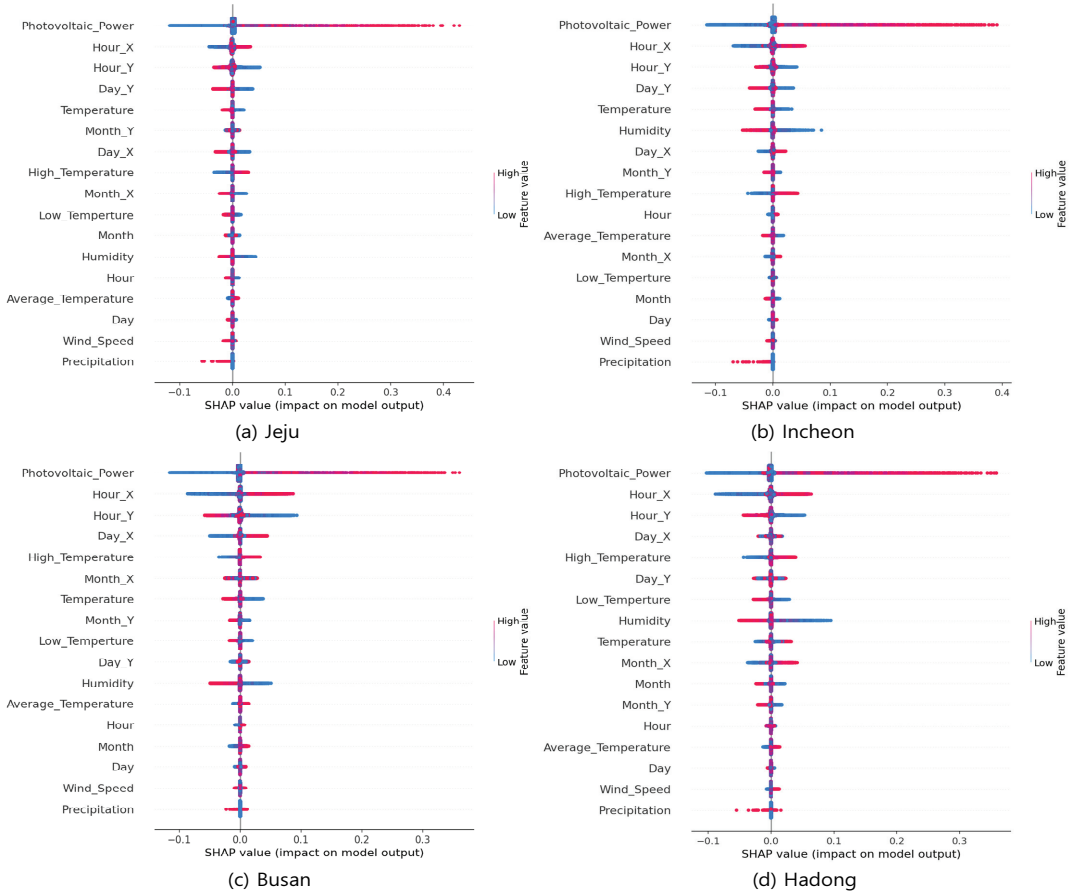


Fig. 4. Summary Plot of Proposed Model

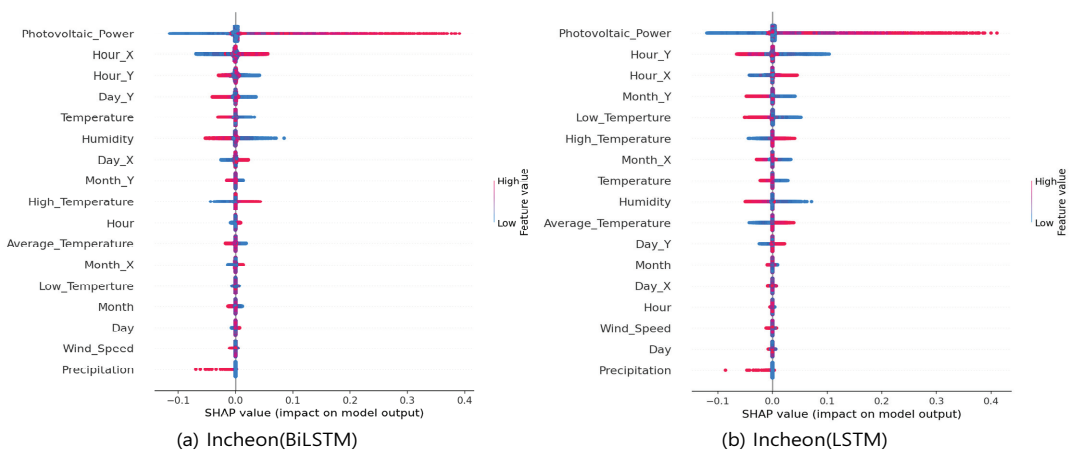


Fig. 5. Summary Plot Comparison Between BiLSTM Model and LSTM Model

큰 영향을 미쳤음을 알 수 있었으며, 기온과 관련된 데이터가 태양광 발전량 예측에 있어서 중요하다는 것을 확인하였다. 또한, BiLSTM 모델의 입력으로 활용되는 과거 데이터가 예측에 매우 큰 영향을 미쳤음을 확인할 수 있었다.

Fig. 5은 인천 지역의 태양광 발전량 데이터를 사용하여 구성한 BiLSTM 모델과 일반적인 LSTM 모델을 SHAP의 summary plot을 통하여 해석한 결과를 나타낸다. Fig. 5의 (a), (b) 모두 시 데이터, 기온과 관련된 데이터, 과거 태양광 발전량 데이터가 중요하다는 것을 확인할 수 있다. 하지만 LSTM 모델에서는 시간대별 기온보다는 해당 날의 최대 기온, 최저 기온의 중요도가 더 높다는 결과를 보여주고 있으며 습도의 중요도가 비교적 낮게 나와있다. 이는 앞에서 언급하였던 LSTM의 단점 때문에 발생하는 것으로 확인된다. 일반적인 LSTM 모델은 바로 직전의 과거 태양광 발전량 데이터에 높은 중요도를 부여하였기 때문에 비교적 낮은 중요도를 갖는 변수들의 영향력에 대하여 잘 학습을 하지 못한다. 반면 BiLSTM은 양방향 정보를 취합함으로써 이러한 문제를 완화시켜줄 수 있다. BiLSTM의 summary plot을 분석해보면 해당 날의 최대 최저 기온보다는 시간대별 기온의 중요도가 더욱 높게 도출되었으며 습도의 중요도 또한 순위가 올라가는 모양을 보여주고 있다. 이는 습도가 높으면 빛의 산란이 많이 일어나 태양광 발전량이 감소할 것이라는 일반적인 상식에 부합한다. 이처럼 BiLSTM은 양방향 정보를 취합함으로써 직전 시점의 태양광 발전량만 집중하는 것이 아닌 다른 외부 요인의 영향력까지 보다 정확하게 학습할 수 있기 때문에 일반적인 LSTM 모델보다 좋은 예측 성능을 보인다고 분석이 가능하다.

6. 결 론

본 논문에서는 BiLSTM을 통해 태양광 발전량을 예측하고 SHAP을 사용하여 예측 결과 도출 과정을 설명하는 예측 기법을 제안하였다. 제주, 인천, 부산, 하동에 위치한 발전소의 1시간 단위의 태양광 발전량 데이터를 대상으로 실험하였다. 다양한 기계학습 및 심층학습 기반의 예측 모델들과 예측 정확도를 비교한 결과, 제안하는 BiLSTM 기반의 예측 모델이 가장 뛰어난 예측 성능을 기록하였다.

또한, 설명가능한 인공지능 모델 중 하나인 SHAP을 통해 태양광 발전량의 예측 결과 도출 과정을 분석하였다. 예측에 활용된 각 입력변수에 대하여 SHAP Value 등을 계산하고, 특징 중요도를 산출하여 분석한 결과, 시간, 기온, 과거 태양광 발전량 데이터가 예측에 큰 영향을 미친 것을 확인할 수 있었다.

일반적인 LSTM과 제안한 BiLSTM 모델을 SHAP을 통해 분석한 결과를 비교하였을 때 LSTM 모델은 시간대별 기온보다는 일 최고 기온, 일 최저 기온의 영향력이 높고 습도의 영향력이 비교적 낮게 도출되는 등 우리의 상식과는 일치하지 않는 결과를 보여주었다. 이는 LSTM 모델이 이전 시점의 태양광 발전량 데이터에 과도하게 집중하였기 때문에 발생한다. 하지만 BiLSTM 모델은 양방향 정보를 취합함으로써 보다 입력변수의 중요도 및 영향력을 우리의 상식에 맞게 잘 분

석하는 것을 확인하였다.

향후, 본 연구에서 제안한 예측 기법을 국내 데이터뿐만 아니라 해외 발전소에서 수집한 다양한 데이터셋에 적용하여 그 효과를 입증할 계획이다. 또한, 예측에 중요한 변수에 집중할 수 있는 attention 메커니즘을 적용한 예측 모델을 구성하여 예측 정확도를 높이고, 데이터 영향력을 보다 심층적으로 분석하여 해외와 같이 다른 기상 환경 조건에서는 어떤 입력 변수가 가장 큰 영향을 주는지 확인할 계획이다.

References

- [1] A. A. A. Abuelnuor, K. M. Saqr, S. A. A. Mohieldein, K. A. Dafallah, M. M. Abdullah, and Y. A. M. Nogoud, "Exergy analysis of Garri '2' 180 MW combined cycle power plant," *Renewable and Sustainable Energy Reviews*, Vol.79, pp.960-969, 2017.
- [2] M. Taylor, P. Ralon, H. Anuta, and S. Al-Zoghoul, "IRENA Renewable Power Generation Costs in 2019," International Renewable Energy Agency, Abu Dhabi, UAE, 2020.
- [3] K. E. Lonngren and E. W. Bai, "On the global warming problem due to carbon dioxide," *Energy Policy*, Vol.36, No.4, pp.1567-1568, 2008.
- [4] S. Shafiee and E. Topal, "When will fossil fuel reserves be diminished?," *Energy Policy*, Vol.37, No.1, pp.181-189, 2009.
- [5] A. Karmaker, M. M. Rahman, M. A. Hossain, and M. R. Ahmed, "Exploration and corrective measures of greenhouse gas emission from fossil fuel power stations for Bangladesh," *Journal of Cleaner Production*, Vol.244, 2020.
- [6] 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.
- [7] H. Chitsaz, H. Shaker, H. Zareipour, D. Wood, and N. Amjady, "Short-term electricity load forecasting of building in microgrids," *Energy and Buildings*, Vol.99, pp.50-60, 2015.
- [8] M. Morandin, S. Bolognani, and A. Faggion, "Active Torque Damping for an ICE-Based Domestic CHP System with an SPM Machine Drive," *IEEE Transactions on Industry Applications*, Vol.51, No.4, pp.3137-3146, 2015.
- [9] A. Khosravi, R. N. N. Koury, L. Machado, and J. J. G. Pabon, "Prediction of hourly solar radiation in abu musa island using machine learning algorithms," *Journal of Cleaner Production*, Vol.176, pp.63-75, 2018.
- [10] M. Lee, W. Lee, and J. Jung, "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.
- [11] B. Zhang, et al., "A multiple time series-based recurrent neural network for short-term load forecasting," *Soft Computing*, Vol.22, No.12, pp.4099-4112, 2018.

[12] J. H. Jeong and Y. T. Chae, "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.

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

[14] 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.

[15] L. Massidda and M. Marrocu, "Use of multilinear adaptive regression splines and numerical weather prediction to forecast the power output of a PV plant in Borkum," *Solar Energy*, Vol.146, pp.141-149, 2017.

[16] L. M. Halabi, S. Mekhilef, and M. Hossain, "Performance evaluation of hybrid adaptive neuro-fuzzy inference system models for predicting monthly global solar radiation," *Applied Energy*, Vol.213, pp.247-261, 2018.

[17] F. A. Kraemer, D. Ammar, A. E. Braten, N. Tamkittikhun, and D. Palma, "Solar energy prediction for constrained IoT nodes based on public weather forecasts," In *Proceedings of the Seventh International Conference on the Internet of Things*, pp.1-8, 2017.

[18] F. Almonacid, C. Rus, P. Perez-Higueras, and L. Hontoria, "Calculation of the energy provided by a PV generator. Comparative study: Conventional methods vs. artificial neural networks," *Energy*, Vol.36, No.1, pp.375-384, 2011.

[19] K. Wang, X. Qi, and H. Liu, "Photovoltaic power forecasting based LSTM-Convolutional Network," *Energy*, Vol.189, pp.116225, 2019.

[20] G. de Freitas Viscondi and S. N. Alves-Souza, "A systematic literature review on big data for solar photovoltaic electricity generation forecasting," *Sustainable Energy Technologies and Assessments*, Vol.31, pp.54-63, 2019.

[21] R. Meenal and A. I. Selvakumar, "Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters," *Renewable Energy*, Vol.121, pp.324-343, 2018.

[22] S. Park, J. Moon, S. Jung, S. Rho, S. W. Baik, and E. Hwang, "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.

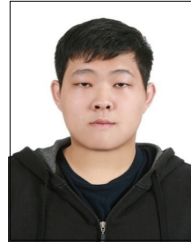
[23] S. Bhanja and A. Das, "Impact of data normalization on deep neural network for time series forecasting," *arXiv Preprint arXiv:1812.05519*, 2018.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, Vol.521, No.7553, pp.436-444, 2015.

[25] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," 2014.

[26] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, Vol.45, No.11, pp.2673-2681, 1997.

[27] S. M. Lundberg and S. I. Lee, "A Unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, Vol.30, pp.4765-4774, 2017.



박 성 우

<https://orcid.org/0000-0001-5469-0332>
 e-mail : psw5574@korea.ac.kr
 2018년 홍익대학교 컴퓨터공학과(학사)
 2018년 ~ 현 재 고려대학교 전기전자공학부
 박사과정
 관심분야 : 인공지능, 빅데이터 분석,
 시계열 예측



정 승 민

<https://orcid.org/0000-0002-4339-4885>
 e-mail : jmkstcom@korea.ac.kr
 2018년 고려대학교 전기전자공학부(학사)
 2018년 ~ 현 재 고려대학교 전기전자공학부
 박사과정
 관심분야 : 인공지능, 데이터 마이닝,
 스마트 그리드



문 재 욱

<https://orcid.org/0000-0002-9835-360X>
 e-mail : jaewookmo@korea.ac.kr
 2019년 고려대학교 전기전자공학부(학사)
 2019년 ~ 현 재 고려대학교 전기전자공학부
 박사과정
 관심분야 : 기계학습, 데이터마이닝, 예측



황 인 준

<https://orcid.org/0000-0002-0418-4092>
 e-mail : ehwang04@korea.ac.kr
 1988년 서울대학교 컴퓨터공학과(학사)
 1990년 서울대학교 컴퓨터공학과(석사)
 1998년 Univ. Maryland at College Park
 전산학과(박사)

1998년 ~ 1999년 Bowie State Univ. 조교수
 1999년 Hughes Research Lab. 연구교수
 1999년 ~ 2004년 아주대학교 정보통신전문대학원 조/부교수
 2004년 ~ 현 재 고려대학교 전기전자공학부 교수
 관심분야 : 데이터베이스, 빅데이터 처리, 영상 처리