

인공지능 알고리즘 기반 미국 Arroyo Grande 시 태양광 발전량 예측 분석

김우현¹, 강민석², 최수현³, 김재현¹, 조석현*
조선대학교¹, 연세대학교², 경기대학교³, *University of California, San Diego (UCSD)

dn001218@gmail.com, k5911839@gmail.com, tngus2255@gmail.com,
kjh01062325507@gmail.com, *justinshcho@gmail.com

Artificial Intelligence-based Predictive Analysis of Photovoltaic Power Generation in Arroyo Grande, CA, USA

Woohyun Kim¹, Minseok Kang², Suhyeon Choi³, Jaehyun Kim¹ and Seokheon Cho *
Chosun University¹, Yonsei University², Kyonggi University³, *University of California, San Diego (UCSD).

Abstract

As abnormal climate related to global warming causes damages frequently around the world, efforts have been being made to develop alternative energy sources that can replace fossil fuels. The demand and importance of solar energy sources among various renewable energy sources have been increasing significantly. Therefore, it is crucial to predict accurate photovoltaic power generation under continuously changing weather conditions for stable and efficient operation of solar power systems. In this paper, we studied the prediction models for photovoltaic power generation based on Artificial Intelligence algorithms using various weather data, which affects solar power generation. The Artificial Intelligence algorithms considered for a photovoltaic power generation prediction model are Multiple Linear Regression, Random Forest Regression, and Multi-Layer Perceptron. In addition, unlike previous studies, we developed a model that predicts solar power generation measured for a short period of 15 minutes. In particular, we identified the different time zones of day for each month when power from solar panels cannot be generated and excluded the data measured from these time zones to obtain a more reasonable performance of the solar power generation prediction model.

I. 서론

오늘날 전 세계적으로 이상기후로 인한 홍수, 가뭄 그리고 광범위한 산불 등 전례 없는 수많은 피해가 늘어나고 있는 상황이다 [1]. 이러한 상황에서의 과도한 화석연료 사용은 더욱 크나큰 지구 온난화와 그에 따른 이상기후 현상을 불러오기에 안정적인 에너지 공급과 환경을 보호하기 위해 대체 에너지원 개발에 힘을 쏟고 있다. 이에 따라 태양광 발전에 많은 관심을 보이고 있으며, 태양광 설치량이 전세계적으로 증가하고 있는 상황이다. 미국의 경우 2023 년 1630 억 kWh 에서 2025 년 2860 억 kWh 로 75% 이상 증가할 것으로 예상된다고 한다 [2]. 하지만 태양광 발전은 일사량이나 구름 량 등의 기상 상황에 따라 변동성이 심하다. 이처럼 불확실성이 큰 태양광 발전은 기존의 전력 발전망의 안정적인 운영을 방해할 뿐만 아니라 불필요한 기저 전력 발전을 야기시킬 수 있다는 문제점들이 존재한다. 따라서 본 연구에서는 안정적인 효율적인 태양광 발전을 위해서 가장 중요한 다양한 기상 상황에 따른 발전량 예측 모델을 제시하고자 한다.

N. Sanewal *et al.*은 총 3 개월간 수집한 일사량과 태양광 모듈 주변 온도만을 독립변수로 사용하여 3 시간 간격의 태양광 발전량을 예측하였다. 이 논문에서는 Linear Regression (LR), Lasso Regression 그리고 Extreme Gradient Boosting (XGBoost) Regression 등 총 6 개의 알고리즘들을 고려하였다 [3]. J. Kaur *et al.*은 4 개월간 수집한 5 분 간격의 온도, 습도, 기압, 풍속, 풍향, 일몰 시간 및 일출 시간을 독립변수로 사용하여

태양광 발전량과 가장 밀접한 일사량을 예측하였다 [4]. LR, Lasso Regression 그리고 Multi-Layer Perceptron (MLP) 등 총 3 개의 알고리즘을 고려하였다. S. Jumaat *et al.*은 20 개월 동안 측정된 시간별 온도, 상대습도, 일사량, 최저온도 그리고 최고온도 등 총 9 개의 기상 데이터를 활용하여 태양광 발전량을 예측하였다 [5]. 전체 데이터는 -1 에서 1 까지의 범위로 정규화를 진행하였고, MLP 기반 두 개의 Artificial Neural Network (ANN) 모델을 고려하였다. J. Kim *et al.*은 2 년 동안 텍사스 주 오스틴 시에서 수집한 1 시간 단위의 기상 데이터를 활용하여 Multi-Linear Regression (MLR) 기반 전력 부하 예보 모델을 제시하였다 [6]. 이 예보 모델은 현재 시점으로부터 1 시간 후 또는 24 시간 후와 같은 단기 전력 부하 예보 모델이다.

위의 선행 연구들은 1 시간 또는 3 시간 간격으로 수집한 기상 데이터들과 태양광 발전량 관련 데이터들을 사용하였다. 이처럼 예측할 태양광 발전량의 시간상 단위가 크면 클수록 일반적으로 예측 모델 성능이 높아지게 된다. 또한, 태양광 발전량 예측 시 고려한 데이터의 양이 3 개월 또는 4 개월처럼 올바르게 예측 모델을 학습시키기에 적었다. 본 연구에서는 이러한 점을 보완하기 위해 2 년 동안 수집한 다양한 기상데이터와 태양 발전량 관련 데이터들을 이용하였다. 또한, 15 분처럼 작은 시간 단위 동안 측정된 태양광 발전량에 대한 예측 모델을 개발하였다. 태양이 없는 대부분의 시간대는 태양광 발전량이 0 이기 때문에, 일광이 없는 시간대의 태양광 발전량을 무의미하게 예측하는 것 대신에 월별로 일광이 없는 시간대를 파악하고 그

시간대의 데이터는 제외시킴으로써 합리적인 태양광 발전량 예측 모델을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 제 II 장에서는 연구에 사용된 데이터 및 전처리에 대해 설명하고, 제 III 장에서는 예측에 사용된 알고리즘에 대한 설명과 분석 및 결과에 대해 설명하였다. 제 IV 장에서는 본 논문의 결과와 향후 연구 과제에 대해 제시하였다.

II. 태양광 발전량 예측을 위한 데이터 및 전처리 과정

2.1 원본 데이터 세트 설명

본 논문에서는 California Distributed Generation Statistics 에서 제공하는 미국 California 주 Arroyo Grande 시에 설치된 하나의 태양광 모듈로부터 수집한 태양광 발전량 데이터들과 SOLCAST 에서 제공하는 위의 태양광 발전 지역에 해당하는 기상 데이터들을 사용하였다 [7, 8]. 표 1 은 상기 데이터들을 포함하고 있는 원본 태양광 발전량 데이터 세트이다. 다음과 같은 독립 변수들을 고려한다. 데이터를 측정된 년, 월, 일과 시간을 포함하는 시점 정보 (date time), 일사량 (solar radiation), 대기의 온도 (air temperature), 이슬점 온도 (dew point temperature), 상대 습도 (relative humidity), 지표면 압력 (surface pressure), 풍속 (wind speed), 풍향 (wind direction), 대기 중 강수 가능 수분량 (precipitable water), 강수량 (precipitation rate), 구름량 (cloud cover), 지표면의 태양광 반사율 (albedo), 북쪽을 기준으로 태양의 위치가 시계 방향으로 몇 도 인지를 나타낸 태양의 방위각 (azimuth) 및 관측지점의 수직선과 태양 사이의 각도인 천정 각 (zenith) 등 총 14 개가 있다. 또한, 고려하는 종속 변수는 태양광 발전량 (solar power)이다. 원본 태양광 발전량 데이터 세트는 2015 년 01 월 01 일부터 2016 년 12 월 31 일까지 2 년간 15 분 간격으로 측정된 데이터들 로서 총 데이터 샘플 개수는 70,175 개이다.

표 1. 원본 데이터 세트의 구성 변수들

Dependent Variables
solar power [wh]
Independent Variables
date time, solar radiation [kw/m ²], air temperature [°C], dew point temperature [°C], relative humidity [%], surface pressure [Pa], wind speed [m/s], wind direction [°], precipitable water [mm], precipitation rate [%], cloud cover [%], albedo [%], azimuth [°], and zenith [°]

2.2 원본 데이터 세트에 대한 전처리 과정

종속 변수인 태양광 발전량의 결측 값들에 대하여 다음과 같이 전처리를 진행하였다. 해가 없는 시간대는 태양광 발전이 불가능하므로 그 시간대에 태양광 발전량에 있어서 결측 치가 있다면 0 으로 대체하였다. 반대로 해가 있는 시간대에서의 태양광 발전량 결측 값들은 선형 보간 법 (Linear Interpolation)을 적용하여 근사한 값들로 복원시켰다.

또한, 독립변수 중 데이터 측정 시점 정보 (date time)를 세분화하여 연도 (year), 월 (month), 일 (day), 시간 (hour) 및 분 (min)처럼 개별 변수로 나누고, 정수형 값 (integer) 형태로 변환시키는 작업을 하였다. 그 중에서 태양광 발전량과 관련성이 가장 높은 특성들로 month 와 hour 두 변수만을 선택하여 사용하였다.

하루 24 시간 데이터를 모두 사용하여 태양광 발전량을 예측하게 되면, 하루 중 거의 절반 정도 되는 해가 뜨지 않은 시간대에 대해서 예측 모델이 당연하게 태양광 발전량을 0 으로 예측하여 모델 성능이 과도하게 높은 현상을 보였다. 이를 방지하기 위해 미국 Arroyo Grande 시에서 월 별로 해가 뜨지 않는 시간대를 파악한 후 해당 시간대의 데이터들을 예측 모델의 고려 대상에서 제외시켰다.

이러한 전처리 과정을 거친 후 본 연구에서 고려할 데이터 세트를 Constructed Feature Dataset (CFD)라고 정의하고, 이 데이터 세트에 포함된 독립 변수들을 표 2 에 나타내었다. CFD 는 15 개의 독립변수를 포함하고 있고 총 데이터 샘플 개수는 41,524 개이다.

표 2. Constructed Feature Dataset (CFD)의 독립 변수

Independent Variables
Month, hour, solar radiation [kw/m ²], air temperature [°C], dew point temperature [°C], relative humidity [%], surface pressure [Pa], wind speed [m/s], wind direction [°], precipitable water [mm], precipitation rate [%], cloud cover [%], albedo [%], azimuth [°], and zenith [°]

III. 태양광 발전량 예측 모델 분석

3.1 인공지능 알고리즘 및 성능 지표

본 연구에서는 Multiple Linear Regression (MLR), Random Forest Regression (RFR) 및 Multi-Layer Perceptron (MLP) 등 총 3 가지 인공지능 알고리즘들을 사용하여 태양광 발전량을 예측하였다. 성능 지표로는 R² (R-squared) 와 Root Mean Squared Error (RMSE)를 사용하였다. 여기에서 R² 는 회귀모델에서 적합도를 판단하는 결정계수로 0 에서 1 사이의 값을 가지게 되고, 1 에 가까울수록 독립 변수들이 종속 변수를 잘 설명한다고 볼 수 있다. RMSE 는 예측 값과 실제 값의 차이를 제공해서 평균한 값의 제공근을 뜻 하고, 이 값이 작을수록 모델의 성능이 높다고 말할 수 있다.

3.2 태양광 발전량 예측 모델 분석 결과

태양광 발전량이 계절에 따른 다르기 때문에 모든 4 계절에 대한 태양광 발전량 예측 모델을 개발하기 위해서, CFD 의 전체 데이터 세트에서 데이터를 2015 년 그리고 2016 년처럼 1 년씩으로 나눈 후 각각 학습용 데이터와 테스트용 데이터로 사용하였다. 특히, 본 연구에서 고려한 RFR 과 MLP 의 최적 하이퍼 파라미터는 다음과 같다. RFR 의 경우 트리 깊이 (tree depth)를 10 이고 모델 개수를 100 이다. MLP 의 경우에는 은닉층의 개수가 40 이고 은닉층당 뉴런의 개수는 40 이며 Epochs 는 500 이다.

그림 1 은 세 개의 알고리즘에 대한 예측 모델 성능 결과인 R² 과 RMSE 를 도시하고 있다. 태양광 발전량을 예측한 결과 세 개 알고리즘들 중에서 MLP 가 R² 값이 0.942 이고 RMSE 값이 65.396 [wh]으로 가장 성능이 좋았다. 짧은 시간 동안 (예: 15 분) 태양광 패널이 생산한 발전량을 예측하는 모델 성능이 긴 시간 동안에 (예: 1 시간 또는 3 시간) 걸친 태양광 발전량 예측 모델 성능보다 일반적으로 떨어짐을 고려하였을 때, 본 연구에서 제시하는 15 분 태양광 발전량 예측 모델에 대한 성능으로 MLP 사용 시 얻게 되는 0.942 의 R² 값은 상당히 높다고 할 수 있다. 또한, 본 논문에서 고려한 2015 년과 2016 년 2 년에 걸쳐 수집한 15 분 동안 측정된 태양광 발전량의 최대값은 899.5 [wh]로 MLP

기본 모델의 RMSE 값 65.396 [wh]와 비교했을 때, 최대 발전량 대비 에러 비율이 7.27 %로 용인될 수 있는 예측 모델 성능이다. 이와는 반대로, MLR 이 R^2 값이 0.848 이고 RMSE 값이 106.105 [wh]로서 가장 성능이 낮았다.

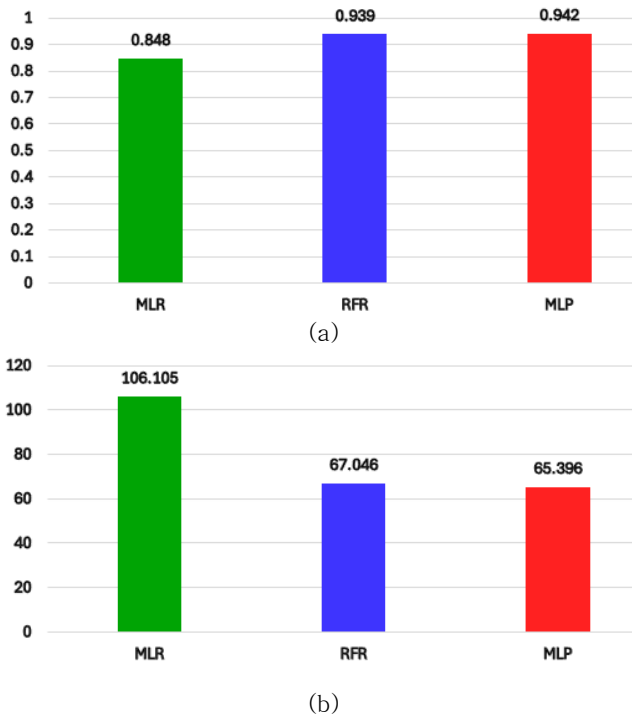


그림 1. 태양광 발전량 예측 모델 별 성능 결과 ((a): R^2 , (b) RMSE [wh])

IV. 결론

본 연구에서는 미국의 California 주 Arroyo Grande 시에 위치한 하나의 태양광 발전기로부터 측정된 15 분 간격의 태양광 발전량 데이터와 SOLCAST 에서 제공한 해당 태양광 발전기가 위한 지역의 기상 데이터를 활용하여 인공지능 알고리즘을 기반으로 한 태양광 발전량을 예측하는 모델을 제안하였다. 고려하는 데이터들에 대한 결측 값 전처리와 예측 모델 성능이 극도로 높아지는 것을 방지하기 위해 Arroyo Grande 시의 월 별 일출 및 일몰 시간들을 파악한 후 일출과 일몰 사이 시간대에 대해서 태양광 발전량 예측을 진행하였다. 또한, 태양광 발전량 예측 모델로서 Multiple Linear Regression (MLR), Random Forest Regression (RFR), Multi-Layer Perceptron (MLP) 알고리즘들을 사용하였다. 예측 모델 성능을 비교하였을 때, MLP, RFR 그리고 MLR 순서대로 좋은 성능을 보였다. 특히, MLP 기반 모델 같은 경우는 R^2 값이 0.960 이고 RMSE 값이 50.517 [wh]로 우수한 성능을 보임을 확인할 수 있었다. 1 시간 또는 3 시간과 같은 다소 긴 시간 동안 측정된 태양광 발전량을 예측한 다른 일반적인 선행 연구들과 달리, 본 연구에서 수행한 짧은 시간인 15 분 간격의 태양광 발전량 예측 분석에서 좋은 성능을 보였다는 점에서 의미가 있다.

태양광 발전 에너지의 보다 안정적인 전력 공급 및 기저 발전원의 전력 생산 감소를 위해서 본 연구 결과를 기반으로 한 단기적 미래 태양광 발전량 예측 모델로 확대하는 향후 연구가 필요하다.

ACKNOWLEDGMENT

Following are results of a study on the "Leaders in Industry–university Cooperation 3.0" Project, supported by the Ministry of Education and National Research Foundation of Korea.

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP (Institute of Information & communications Technology Planning & Evaluation). (No.2021-0-01393) & (No.2023-0-00054).

참 고 문 헌

- [1] H. Fountain, "Climate Change Affected Australia's Wildfires, Scientists Confirm," The New York Times. Mar. 2020.
- [2] K. Antonio, "Solar and wind to lead growth of U.S. power generation for the next two years," U.S. Energy Information Administration (EIA), Jan. 2024.
- [3] N. Sanewal and V. Khanna, "Solar Power Prediction in North India Using Different Regression Models," 2023 IEEE World Conference on Applied Intelligence and Computing (AIC), pp. 1-6, Oct. 2023.
- [4] J. Kaur, A. Goyal, P. Handa, and N. Goel, "Solar power forecasting using ordinary least square based regression algorithms," 2022 IEEE Delhi Section Conference (DELCON), pp. 1-6, Apr. 2022.
- [5] S. Jumaat, F. Crocker, M. Wahab, N. Radzi, and M. Othman, "Prediction of Photovoltaic (PV) Output Using Artificial Neural Network (ANN) Based on Ambient Factors," Journal of Physics: Conference Series, Dec. 2017.
- [6] J. Kim, S. Cho, K. Ko, and R. R. Rao, "Short-term Electric Load Prediction Using Multiple Linear Regression Method," 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), pp. 1-6, Dec. 2018.
- [7] California Distributed Generation Statistics, "CSI 15-Minute Interval data," 2010.
- [8] SOLCAST a DNV company, "Solcast Historical Time Series," 2007.