

강건한 태양광 발전량 예측을 위한 2단계 신경망 최적화

오진영*, 소다영**, 문지훈*^o
*순천향대학교 AI·빅데이터학과,
**순천향대학교 ICT융합학과,
^o순천향대학교 ICT융합학과

e-mail: {wlsdud8261, sodayeong, jmoon22}@sch.ac.kr

Two-Stage Neural Network Optimization for Robust Solar Photovoltaic Forecasting

Jinyeong Oh*, Dayeong So**, Jihoon Moon*^o

*Dept. of AI and Big Data, Soonchunhyang University,

**Dept. of ICT Convergence, Soonchunhyang University,

^oDept. of ICT Convergence, Soonchunhyang University

● 요약 ●

태양광 에너지는 탄소 중립 이행을 위한 주요 방안으로 많은 주목을 받고 있다. 태양광 발전량은 여러 환경적 요인에 따라 크게 달라질 수 있으므로, 정확한 발전량 예측은 전력 네트워크의 안정성과 효율적인 에너지 관리에 근본적으로 중요하다. 대표적인 인공지능 기술인 신경망(Neural Network)은 불안정한 환경 변수와 복잡한 상호작용을 효과적으로 학습할 수 있어 태양광 발전량 예측에서 우수한 성능을 도출하였다. 하지만, 신경망은 모델의 구조나 초매개변수(Hyperparameter)를 최적화하는 것은 복잡하고 시간이 많이 드는 작업이므로, 에너지 분야에서 실제 산업 적용에 한계가 존재한다. 본 논문은 2단계 신경망 최적화를 통한 태양광 발전량 예측 기법을 제안한다. 먼저, 태양광 발전량 데이터 셋을 훈련 집합과 평가 집합으로 분할한다. 훈련 집합에서, 각기 다른 은닉층의 개수로 구성된 여러 신경망 모델을 구성하고, 모델별로 Optuna를 적용하여 최적의 초매개변수값을 선정한다. 다음으로, 은닉층별 최적화된 신경망 모델을 이용해 훈련과 평가 집합에서는 각각 5겹 교차검증을 적용한 발전량 추정값과 예측값을 출력한다. 마지막으로, 스택킹 앙상블 방식을 채택해 기본 초매개변수값으로 설정해도 우수한 성능을 도출하는 랜덤 포레스트를 이용하여 추정값을 학습하고, 평가 집합의 예측값을 입력으로 받아 최종 태양광 발전량을 예측한다. 인천 지역으로 실험한 결과, 제안한 방식은 모델링이 간편할 뿐만 아니라 여러 신경망 모델보다 우수한 예측 성능을 도출하였으며, 이를 바탕으로 국내 에너지 산업에 이바지할 수 있을 것으로 기대한다.

키워드: 태양광 발전량 예측(solar photovoltaic forecasting), 신경망(neural network), 초매개변수 최적화(hyperparameter optimization) 스택킹 앙상블(stacking ensemble),

I. Introduction

최근 지구 온난화 및 환경 문제에 따른 대응책으로, 재생 에너지의 중요성이 강조되고 있다. 태양광(Photovoltaic)은 특히, 무공해, 지속/재생 가능한 자원이라는 특성으로 기후 변화 문제에 대한 해결책으로 많은 주목을 받고 있다[1]. 또한, 태양광 발전은 설치의 용이성, 무한한 자원, 친환경적 특성뿐만 아니라 다른 재생 에너지원과 비교하면 소음이나 오염물질 배출 없이 발전이 이루어진다는 장점이 존재한다. 이러한 특성으로 인해 향후 태양광 에너지의 발전과 활용 비중이

지속해서 증가할 것으로 보인다[2].

태양광 발전량 예측은 전력 생산과 소비 간의 균형 유지, 최적의 에너지 자원 활용 등 지속 가능한 에너지 시스템 운영에 매우 중요하다. 하지만, 태양광 발전은 다른 에너지 자원들과 비교할 때 환경 요소에 완전히 의존하며, 시간에 따라 동적으로 변하는 특성이 있으므로, 태양광 발전량을 정확하게 예측하는 것은 매우 어려운 문제이다[3]. 이러한 복잡한 상황을 고려하여, 인공지능 기법은 환경적 불확실성

속에서도 발전량을 정확하게 예측하는 도구로, 그 필요성이 점점 강조되고 있다[4].

초기 연구들은 주로 ARIMA(Autoregressive Integrated Moving Average)와 같은 전통적인 시계열 예측 기법들을 활용하여 태양광 발전량 예측에 접근하였다[5, 6]. 하지만 환경 변수가 가진 복잡한 상관관계와 비선형성으로 인해 이러한 기법들로는 정확한 태양광 발전량 예측이 어려웠다. 이러한 한계를 극복하고자 최근에는 인공지능 기술이 예측 모델링에 적용되고 있다. 특히, 신경망 모델은 구조와 학습 방식이 복잡한 패턴의 데이터를 처리하는 데 탁월하다는 장점으로 많은 주목을 받고 있다[7, 8].

하지만 신경망은 다양한 초매개변수(Hyperparameter)를 포함하고 있어, 모델 학습에서 최적화 과정이 복잡하다[9]. 이로 인해 인공지능 관련 지식이 다소 부족한 에너지 분야에서의 실제 적용에 제약이 발생할 수 있다. 이러한 한계를 극복하고자, 본 논문은 간단한 태양광 발전량 예측을 위한 2단계 신경망 모델의 초매개변수 최적화 기법을 제안한다. 1단계에서는 Optuna를 이용해 여러 신경망 모델의 초매개변수를 최적화한다. 2단계에서는 딥러닝 모델의 과적합을 방지하고자 스택킹 앙상블 방법을 도입하여 강건한 태양광 발전량 예측 체계를 구축한다.

본 논문의 나머지 구성은 다음과 같다. 2장은 데이터 셋 구성과 전처리 과정에 대해 설명한다. 3장에서 본 연구에 적용된 최적화 방법과 실험 과정을 상세히 다룬다. 4장과 5장은 제안한 기법의 효용성을 입증하기 위한 실험 결과와 본 연구의 결론 및 향후 연구 방향을 제시한다.

II. Data Collection and Preprocessing

1. Data Collection

본 논문은 국내 에너지 산업의 적용 가능성을 높이고자 인천 지역 내의 태양광 발전 시설 중 신인천전망대 발전 시설에서 2018년 1월부터 2019년 12월까지 1시간 단위로 수집된 태양광 발전량 데이터를 사용하였다. 또한, 태양광 발전량의 환경적 요인으로 기상청 기상자료 개방포털에서 같은 기간의 시간대별 기온과 해당일의 평균, 최저, 최고 기온 및 습도, 풍속, 강수량의 데이터를 수집하였다.

2. Data Preprocessing

태양광 발전량은 시간과 계절의 변화에 가장 큰 영향을 받는다. 그러나 일몰부터 일출까지의 시간대의 경우 어떠한 기상 값과 관계없이 0이 기록되므로, 본 연구는 08시부터 18시까지를 고려하였다. 시간(Time)의 주기성을 반영하기 위하여 월, 일 변수를 2차원으로 변환하였다[9]. 이와 달리 시간(Hour) 변수는 기존의 24개(0-23)에서 11개(8~18)로 축소되어 2차원으로 변환하였을 때 주기성(Periodicity)을 반영하기에 어려움이 있으므로, 원-핫 인코딩(One-Hot Encoding)을 통해 11개의 변수를 생성하였다. 모델링을 위해 이전에 수집한 기상 변수들과 함께 최소-최대 정규화(Min-Max Normalization)를 수행하였다. 표 1은 본 연구에서 사용된 24개의

입력 변수의 정보를 나타낸다.

Table 1. List of Input Variables

No.	Variables	No.	Variables	No.	Variables
1	Month _{sin} (Numeric)	9	10 a.m. (Binary)	17	6 p.m. (Binary)
2	Month _{cos} (Numeric)	10	11 a.m. (Binary)	18	Temp _{Avg} (Binary)
3	Day _{sin} (Numeric)	11	12 p.m. (Binary)	19	Temp _{Min} (Binary)
4	Day _{cos} (Numeric)	12	1 p.m. (Binary)	20	Temp _{Max} (Binary)
5	6 a.m. (Binary)	13	2 p.m. (Binary)	21	Temp (Binary)
6	7 a.m. (Binary)	14	3 p.m. (Binary)	22	Humi (Binary)
7	8 a.m. (Binary)	15	4 p.m. (Binary)	23	Wind speed (Binary)
8	9 a.m. (Binary)	16	5 p.m. (Binary)	24	Precipitation (Binary)

III. Two-Stage Neural Network Optimization

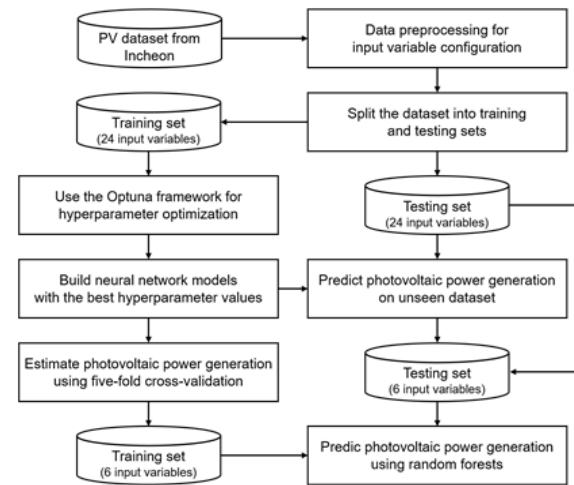


Fig. 1. System Architecture

그림 1은 제안하는 태양광 발전량 예측 방법의 전체적인 아키텍처를 보여준다. 먼저, 전처리된 데이터 셋을 5:5의 비율로 훈련 집합(2018년)과 평가 집합(2019년)으로 분할한다. 훈련 집합에서 6개의 심층 신경망 모델을 대상으로 2단계의 최적화 기법을 수행하고 평가 집합에 최종 예측을 수행한다.

1. 1단계: Optuna

본 논문은 1단계 최적화 방법으로, 초매개변수 최적화를 위한 오픈 소스 라이브러리인 Optuna [10]를 적용한다. Optuna는

Bayesian과 Gradient 기반 최적화 방법론들과 Evolutionary 알고리즘 등과 같은 다양한 최적화 전략을 통합한 기법이다. 특히 Optuna는 동적인 초매개변수 공간(Dynamic Hyperparameter Space)과 조건부 초매개변수(Conditional Hyperparameters)를 지원한다. 해당 방법론은 모델 학습에서 초매개변수의 범위와 구조를 동적으로 조절하여 모델 성능 향상에 중점을 둔다. 그리하여 모델 설정에 필요한 시간과 계산 비용을 줄이는 효과와 예측 성능 모두 개선되는 잠재적인 이점을 제공한다. 따라서 심층 신경망 모델과 같이 복잡한 기계학습 모델 구조에 매우 적합하다.

본 논문에서는 훈련 집합에서 은닉층이 2개부터 7개까지 구성된 6개의 심층 신경망을 고려하여 5겹 교차검증을 통해 신경망별로 최적의 초매개변수값을 선정 및 이를 적용하여 훈련 집합은 5겹 교차검증을 통한 태양광 발전량 추정값, 평가 집합은 태양광 발전량 예측값을 생성하였다.

2. 2단계: 스택킹 앙상블

스택킹 앙상블(Stacking Ensemble)은 다양한 기본 학습 모델의 예측값을 새로운 학습 모델의 입력으로 사용하여 최종 예측값을 생성한다[9]. 앞서 구성한 심층 신경망들은 데이터의 복잡한 패턴을 포착할 수 있으나, 동시에 과적합 될 수 있다는 문제가 있다[9]. 스택킹 앙상블은 기본 학습 모델들의 각각의 강점을 취하고 약점을 보완하여 강력한 성능을 달성할 수 있으므로, 모델의 과적합 위험을 줄이고 서로 다른 모델의 예측 성능 간의 균형을 이룰 수 있다.

랜덤 포레스트는 여러 의사결정 나무를 결합한 앙상블 기법으로 [11], 과적합을 방지하고 다중공선성 문제에 대한 저항성을 지닌다. 이로 인해 스택킹 앙상블의 두 번째 단계 모델로 적합하다. 본 연구는 훈련 집합에서 최적의 신경망 모델들의 추정값으로 구성된 6개의 입력 변수를 학습하여 평가 집합에서의 신경망 모델의 예측값을 입력으로 받아 최종 태양광 발전량을 예측하였다. 이를 바탕으로 신경망과 랜덤 포레스트의 강점을 결합하여 태양광 발전량 예측의 정확도를 극대화할 수 있었다.

IV. Experimental Results

본 연구의 모든 실험은 Python 환경에서 수행하였으며, scikit-learn의 MLPRegressor를 통해 신경망 모델을 구성하였다. LSTM, TCN 등과 같은 복잡한 신경망 모델은 GPU에서 연산해야 하지만, 본 연구는 제한된 여건으로 인해 CPU 환경에서 성능 최적화에 중점을 두었다.

시계열 예측 문제에서 활용되는 의사결정 나무 기반의 기계학습 모델들과 단일 신경망 모델들을 기준으로 예측 성능을 비교하였다. 평가 지표로는 평균 절대 오차(Mean Absolute Error, MAE)와 평균 제곱근 오차(Root Mean Square Error, RMSE)를 사용하였다. 표 2에 평가 지표에 따른 모델 간의 예측 성능을 비교한 결과를 나타내었으며, 제안한 방법은 기계학습 모델 및 단일 신경망 모델들보다 우수한 예측 결과를 도출하였음을 확인할 수 있었다.

Table 2. Prediction Performance Comparison

Models	MAE	RMSE
Decision Tree	158.886	226.601
Random Forest	158.476	225.874
LightGBM	146.240	207.653
DNN _{HL=2}	146.445	210.043
DNN _{HL=3}	146.246	208.147
DNN _{HL=4}	147.960	212.697
DNN _{HL=5}	147.960	212.697
DNN _{HL=6}	148.904	212.864
DNN _{HL=7}	148.233	214.357
Ours	140.823	199.135

V. Conclusions

본 연구는 인천 지역의 태양광 발전 시설 데이터를 활용하여 태양광 발전량을 예측하기 위한 2단계 신경망 최적화 기법을 제안하였다. 제안한 방법론의 성능 평가 결과, 기존 기계학습 모델들보다 상당한 우위를 보였으며, 초매개변수 최적화를 통해 연구 과정에서의 시행착오를 최소화하면서 안정적이고 정확한 예측 성능을 도출하였다.

향후, 본 연구의 방법론을 국내 다양한 지역에 확장하여 태양광 발전량 예측의 일반화를 추구할 예정이다. 현재의 연구 환경이 CPU 중심이었던 점을 고려하여, 향후 GPU 환경에서 고도화된 딥러닝 모델 최적화를 수행하여 태양광 발전량 예측의 정확도를 더욱 향상함으로써 연구 범위를 확장하고자 한다.

ACKNOWLEDGEMENT

본 연구는 순천향대학교 연구지원 사업 및 한국연구재단 4단계 두뇌한국 21 사업(4단계 BK21 사업)의 지원을 받아 작성되었음(과제번호: 5199990514663).

REFERENCES

- [1] P. Chen, S. Liu, X. Wang, and I. Kamwa, "Physics-Shielded Multi-Agent Deep Reinforcement Learning for Safe Active Voltage Control With Photovoltaic/Battery Energy Storage Systems," *IEEE Trans. Smart Grid*, vol. 14, no. 4, pp. 2656-2667, July 2023.
- [2] "Diagnosis of Solar Power Generation Equipment Failures Without On-Site Visits," *Donga Science*, 2023. [Online]. Available: <https://www.dongascience.com/news.php?id=61712>. [Accessed: 20-09-2023].

- [3] K. Hu, S. Cao, L. Wang, W. Li, and M. Lv, "A new ultra-short-term photovoltaic power prediction model based on ground-based cloud images," *J. Cleaner Prod.*, vol. 200, pp. 731-745, 2018.
- [4] S. Park, S. Jung, J. Moon, and E. Hwang, "Explainable Photovoltaic Power Forecasting Scheme Using BiLSTM," *KIPS Trans. Softw. Data Eng.*, vol. 11, pp. 339-346, 2022.
- [5] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Appl. Energy*, vol. 259, article no. 114216, 2020.
- [6] S. Wang, D. Li, G. Liu, Z. Ling, and D. Wang, "Short-term PV Power Prediction Based on Bi-directional Gated Recurrent Unit Network and Adaptive Chirp Mode Decomposition," in *Proc. 2023 3rd Int. Conf. Neural Networks, Information and Communication Engineering (NNICE)*, pp. 702-706, 2023.
- [7] J. Liang and W. Tang, "Ultra-short-term spatiotemporal forecasting of renewable resources: An attention temporal convolutional network-based approach," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 3798-3812f, 2022.