

국지예보모델 및 지상 관측자료를 활용한 머신러닝 기반 일사량 예측 정확도 개선에 관한 연구

김승호, 박상호, 최순호, 반민호, 백자현

한국전력공사 전력연구원

seungho_kim@kepcoco.kr, alegole73@kepcoco.kr, soonhochoi@kepcoco.kr,

minho.ban@kepcoco.kr, jahyun.baek@kepcoco.kr

A Study on improving the accuracy of insolation prediction based on machine learning using LDAPS and ASOS data

Kim Seungho, Park Sangho, Choi Soonho, Ban Minho, Baek Ja-hyun

KEPCO Research Institute.

요약

온실가스 배출량 감소에 관한 관심과 늘어나는 전력 수요를 대처하기 위해 신재생에너지 발전량 예측에 대한 중요성은 갈수록 커지고 있다. 본 논문에서는 신재생발전원 중 태양광 예측에 기반이 되는 일사량 데이터를 기존의 예보보다 정확도를 높일 수 있는 모델을 제시하고자 한다. 국지예보모델(Local Data Assimilation and Prediction System: LDAPS)의 일사량 예보데이터의 정확도 개선을 위해 지상 관측자료를 함께 활용하여 머신러닝 기반 학습을 통해 예측 정확도를 개선하였다. 국지예보모델 데이터에서 지상 관측자료의 일사량 데이터에 영향을 줄 수 있는 데이터를 선정하여 활용하였고, 순환 신경망 모델을 활용한 머신러닝의 학습 이전과 이후 데이터의 분석을 통해 예측 정확도 개선을 설명한다.

I. 서론

전 세계적으로 온실가스 감축을 위해 신재생에너지에 관련한 기술 발전에 큰 노력을 하고 있고, 이는 온실가스 배출량을 줄이는 데 획기적인 대안이 될 수 있다. 2050 탄소중립 시나리오에 따르면 2050년 화력발전을 전면 중단하고 신재생에너지를 확대하여 전환 분야의 순 배출량 0을 목표로 하고 있다. 이러한 상황에서 태양광과 바람으로부터 전기를 생산하는 태양광 발전과 풍력발전 기술은 탄소 배출 저감에 필수적인 기술이다.[1]

제9차 전력 수급 기본계획에 따르면 2030년 재생에너지 발전량 125.7TWh 중 태양광은 45.5TWh(36%), 풍력발전은 40.1TWh(32%)의 비중을 차지하고 있는 등 재생에너지는 지속적으로 확대되는 추세이다.[2] 하지만 신재생에너지 발전은 기후 변화에 따라 변하는 특성이 있으며, 이러한 특성은 전력계통의 불확실성을 증가시켜 안정적인 계통운전을 어렵게 한다. 신재생에너지 활용을 극대화하기 위해서 신재생발전량 예측의 중요성이 주목받기 시작했다.[3]

신재생발전원 중에서 태양광 발전이 큰 비중을 차지하고 있다. 태양광 발전은 기상 데이터 중에서 일사량과 가장 밀접한 관련이 있어 정확한 일사량을 예측하게 되면 태양광 발전 예측의 정확도 향상에 기여할 수 있다.[4] 기상청에서는 기상이나 기후 예측을 위해 수많은 기상 데이터와 물리·역학 방정식을 활용한다. 그중에서 수치예보모델은 관측자료를 가지고 수치적인 계산을 통해 지구를 바둑판과 같은 격자로 나누어 격자점마다 기상과 기후를 예측하는 것이다. 수치예보는 관측기술, 초고속 통신 기술, 자료동화 기술, 슈퍼컴퓨팅 기술 등과 같은 첨단 기술들의 융합을 통해서 이루어진다.[5] 수치예보모델은 현재 여러 가지 모델로 운영되고 있으며, 본 논문에서는 기상청의 국지예보모델을 이용하였다. 국지예보모델은 약 1.5km의 공간 해상도를 가지고 일 4회 48시간까지의 한반도 날씨

예측데이터를 제공한다. 기상청에서는 기상이나 기후의 예측뿐만 아니라 다양한 기상 장비를 활용하여 실시간으로 기상을 관측하고 지상관측 데이터로 정보를 제공하고 있다.

정확한 재생에너지 발전 예측을 위해서는 정확한 기상예보데이터가 필요하지만 실제로 관측자료와 비교해보면 오차가 발생하여 정확도 개선이 필요하다. 따라서, 기상예보와 관측데이터 간 머신러닝을 통해 기존의 기상예보와 관측데이터 간 정확도보다 학습 이후 결과의 정확도를 더 개선하는 것을 목표로 한다.

II. 본론

2.1 데이터 취득

본 논문에서는 전라북도 고창군 인근 지상 관측소의 위치에서 KST(Korea Standard Time) 기준 2022년 7월 1일 0시부터 10월 31일 23시까지 1시간 단위로 수집된 데이터를 이용하였다.

국지예보모델에서는 예보발표시각 0시(UTC) 기준으로 데이터를 KST로 환산하여 사용하였고, 그 외에도 일사량(W/m²), 습도(%), 운량(%)이 포함된 데이터를 이용하였다. 지상 관측자료에서는 관측시각(KST), 일사량(MJ/m²)이 포함된 데이터를 이용하였다. 예측과 관측데이터에서는 일사량 데이터 단위 통일을 위해 국지예보모델 기준인 (W/m²)으로 환산하여 데이터를 이용하였다.

2.2 데이터 전처리 및 학습 모델 생성

2.2.1 데이터 전처리

정확한 머신러닝 모델을 생성하기 위해서는 입력 데이터의 전처리 과정이 필수적이다. 결측치가 존재하는 경우에는 머신러닝 결과에 큰 영향

을 미칠 수 있어 결측치에 대한 처리가 필요하다. 지상 관측자료는 관측소의 상황에 따라 누락된 데이터가 발생할 수 있어 결측치가 존재하는 일자에는 데이터를 제거하였다.

머신러닝에서 모델의 정확도를 높이기 위해서 정규화 기법을 이용한다. 각 변수가 가지는 단위나 값들의 분포가 일정하지 않기 때문에 이러한 영향을 반영하기 위해 평균이 0이고 분산이 1인 정규 분포를 가지게 하는 표준화 기법(Standard Scaler)을 사용하였다.

2.2.2 예측 모델

시계열 데이터의 예측 모델에 쓰이는 순환 신경망(Recurrent Neural Network) 모델인 LSTM(Long short-term memory)를 사용하였다. LSTM 모델은 직전 데이터뿐만 아니라 더 이전의 과거 데이터를 고려하여 앞선 미래의 데이터를 예측하는 데에 쓰인다. 전체 데이터에서 약 8:2 비율로 훈련 데이터셋과 테스트 데이터셋으로 구분하였다. 입력 데이터로 국지예보모델의 일사량, 습도, 운량을 사용했고, 실적 데이터로는 지상관측 데이터의 일사량 데이터를 사용했다.

2.3 데이터 분석

예측 성능을 검증하기 위해서 평가 지표로 MAE(Mean Absolute Error), RMSE(Root Mean Square Error), 결정계수(R2 score)를 이용하였다. MAE와 RMSE는 값이 작을수록 오차가 적은 것이고, 결정계수의 경우에는 1에 가까울수록 데이터가 비슷함을 의미한다. 또한, 야간에는 일사량이 0이기 때문에 실질적인 예측 활용 결과 비교를 위해서 일사량 최댓값의 10% 이상에 해당하는 데이터 검증 결과들을 표 1에 정리하였다.

표 1에서 학습 이전과 이후의 평가 지표를 비교했을 때, MAE는 22.12857에서 21.61701로 줄었고 RMSE는 53.84541에서 45.46418로 값이 더 줄어들었다. 결정계수는 0.95403에서 0.96805로 증가하는 것으로 보아 학습 후의 평가 지표가 우수한 것을 확인할 수 있다. 일사량 최댓값의 10% 이상 데이터들의 평가 지표를 비교해보더라도 학습 후 데이터의 결과가 우수한 것을 볼 수 있다.

평가 지표	학습 전	학습 후
MAE	22.12857	21.61701
RMSE	53.84541	45.46418
R2	0.95403	0.96805
MAE(10%)	54.57676	53.95654
RMSE(10%)	87.00414	73.58499
R2(10%)	0.78350	0.84840

표 1. 학습 모델의 평가 지표

그림 1은 국지예보모델(LDAPS)과 지상관측(ASOS)의 일사량 데이터를 그래프로 나타낸 것이고, 그림 2는 지상관측 일사량과 머신러닝 모델의 예측 결과를 그래프로 나타낸 것이다. 그래프를 비교했을 때 맑은 날에는 국지예보모델도 관측데이터와 정확한 성능을 보였지만 날씨가 흐린 날에는 예보데이터가 큰 오차를 보인다. 하지만 머신러닝을 통해 예측한 그래프를 보면 흐린 날에 데이터가 더 정확하게 예측되는 것을 확인할 수 있다.

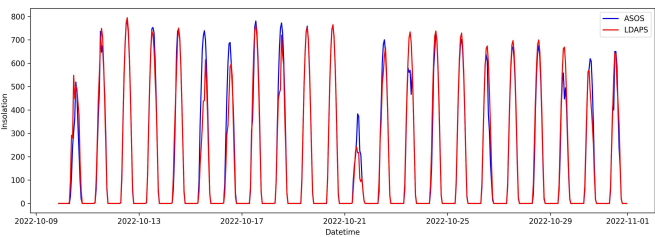


그림 1. 국지예보모델(LDAPS)과 지상관측(ASOS)의 일사량

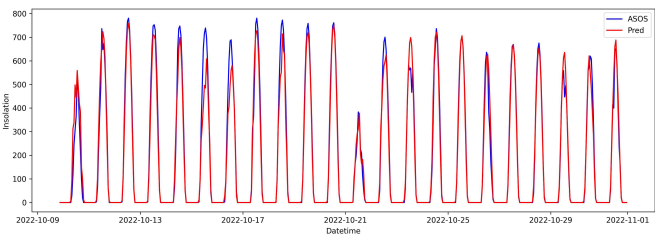


그림 2. 지상관측(ASOS) 일사량과 예측값

III. 결 론

본 논문에서는 국지예보모델과 지상관측 데이터를 머신러닝 학습을 통해 예보데이터의 정확도를 개선하였다. 순환 신경망 모델의 LSTM을 활용하여 예측 모델을 설계했고, 순환 신경망 모델을 활용했을 때 데이터가 더 우수한 성능을 보였다.

향후 연구에서는 더 정확한 예측을 위해 국지예보모델에서 일사량 데이터의 상관계수를 비교하여 관련성이 높은 데이터를 추가하여 학습하고자 한다. 또한, 제공된 기상 데이터뿐만 아니라 위성 데이터를 가지고 구름의 움직임 정보도 일사량 예측 모델에 반영하고자 한다. 궁극적으로 정확도가 개선된 일사량 데이터를 활용하여 태양광 발전량 예측 모델설계에 연계하여 연구를 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 한국전력공사 전력연구원의 “신재생에너지 유연접속기술 개발 및 출력제어기술 실증연구(R22TA25)” 과제의 지원을 받아 작성된 논문임.

참 고 문 헌

[1] 관계부처 합동, 2021, 2050 탄소중립시나리오안.
[2] 산업통상자원부, 2020, 제9차 전력수급기본계획(2020~2034).
[3] A. Hammer et al., "Solar energy assessment using remote sensing technologies," Remote Sensing of Environment, 86 (2003) pp. 423-432.
[4] D. Ryu et al., "Optimum MPPT Control Period for Actual Insolation Condition," 2018 IEEE International Telecommunications Energy Conference, 24 (2019) pp. 1-4.
[5] A. C. Lorenc, "Analysis methods for numerical weather prediction," Quart. J. R. Met. Soc., 112 (1986) pp. 1177-1194.