

기상 패턴 기반 태양광 발전량 예측 모델 선택 방법

강동윤, 박강구
주식회사 케이티

{dongyoon.kang, kanggu.park} @kt.com

Meteorological pattern-based model selection of photovoltaic forecasting

Dongyoon Kang, Kanggu Park
KT Corporation

요 약

전력 계통의 원활한 운영 및 계획을 위해 재생 에너지 분야의 발전량 예측은 매우 중요하다. 그러나 기상 변화, 에너지 생산 시스템의 복잡성, 지리적인 변동성 등의 요인으로 인해 정확한 예측이 어렵다. 단일 모델로 예측할 경우, 모델마다 특정 기상 패턴에 대해 성능 차이가 존재하기 때문에 모든 기상 패턴에 대해 유동적으로 대응할 수 없다는 단점이 있다. 본 논문에서는 발전량 데이터를 이용한 시계열 모델인 EWM 모델, 기상 변수를 사용한 딥러닝 기반 weather, solar 모델 중 예측일의 기상 패턴을 고려하여 태양광 발전량 예측에 사용할 최적 모델을 선택하는 로직을 제안한다. 해당 로직을 적용했을 때 단일 모델보다 예측 성능과 안정성이 더 우수한 것을 확인할 수 있다.

I. 서 론

현재 범지구적인 기후 변화에 대한 위험성이 대두되고 있고, 이에 대응하여 온실가스 배출 규제 및 친환경에너지 사용에 대한 요구가 증가하고 있다. 전 세계의 주요 기업들은 기업이 사용하는 전력의 100%를 태양광과 풍력 등의 재생에너지를 이용해 생산된 전력으로 조달하는 RE100(Renewable Energy 100) 캠페인에 참여하고 있다. 하지만 재생에너지는 기상조건에 따라 발전량이 크게 좌우되기 때문에 전력계통의 간헐성 해소 및 안정화를 위해서는 정확한 발전량 예측이 필요하다. 우리나라에서는 ‘발전량 예측 인센티브 제도’를 통해 발전량 예측의 성능 향상을 독려하고 있고, 이로 인해 발전량 예측 정확도는 수익에 직접적인 영향을 미친다. [1]과 같이 태양광 발전량 예측을 위한 다양한 알고리즘과 모델들이 제안되었다. 하지만 각 모델에는 저마다 한계와 제약 사항이 존재하기 때문에 실제 운영 시의 예측 정확도와 신뢰성을 보장하기 어렵다. 따라서 기존 연구의 한계를 극복하고 개선된 태양광 발전량 예측 로직을 제시할 필요가 있다.

태양광 발전량 예측 모델은 사용하는 기상 변수에 따라 특정 기상 조건에 대한 장단점이 존재한다. [2]와 같이 기상 조건에 따른 모델의 예측 성능의 차이에 대한 연구도 이뤄지고 있다. 본 논문에서는 [3] 연구에서 제시한 모델 선택 및 최적화 전략을 참고하여 각 모델에 대한 기상 조건에 따른 성능 차이 분석과 일별 최적 모델 선택 로직을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 서로 다른 변수를 사용한 3가지 모델을 소개하고, 각 모델의 기상 조건에 따른 성능 차이를 분석한다. 이를 고려한 일별 최적 모델 선택 로직을 소개하고, 실험을 통해 결과를 분석한다.

II. 본론

II.1. 태양광 발전량 예측 모델

본 논문에서는 서로 다른 변수를 이용한 3가지 모델을 사용한다. 1) 기상 변수를 사용하지 않고 최근 발전량 데이터에 지수가중평균을 적용하여 예측값을 생성하는 exponentially weighted moving-average (EWM) 모델, 2) 기온, 습도, 날씨코드(맑음, 구름많음, 흐림, 비/눈)를 입력변수로 사용하는 weather 모델, 3) 일사량, 운량, 온도, 습도 등을 입력변수로 사용하는 solar 모델이다. 일반적인 EWM 모델은 식 (1)과 같이 계산된다.

$$\hat{Y}_t = \begin{cases} Y_1, & t = 1 \\ \alpha \cdot Y_t + (1 - \alpha) \cdot \hat{Y}_t, & t > 1 \end{cases} \quad (1)$$

여기서 α 는 평활 계수로써, $0 < \alpha \leq 1$ 이다. weather 모델과 solar 모델은 RNN 계열의 딥러닝 모델로 구현했다. 일반적인 시계열 모델은 선형적인 추세를 가정하고 예측하기 때문에 비선형적인 패턴이나 변동을 예측할 때는 다층 신경망을 이용해 학습하는 딥러닝 모델보다 예측 성능이 떨어진다는 한계가 있다. 그러나 비슷한 날씨가 지속될 경우에는 오히려 시계열 모델의 성능이 더 뛰어날 수 있어, 기상 조건에 따라 적절한 모델을 선택하는 것이 필요하다.

II.2. 최적 모델 선택 로직

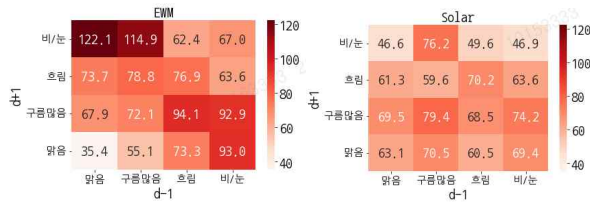
II.2.1. 기상 패턴에 따른 모델별 성능 분석

모델별 성능을 평가할 때, 성능은 시간대별 평균 절대 오차(MAE)로 평가하며 식 (2)와 같이 계산된다.

$$MAE = \frac{1}{24} \sum_{t=1}^{24} |Gen_t - Exp_t| \quad (2)$$

여기서 Gen_t 는 발전소의 시간대별 발전량, Exp_t 는 시간대별 예측값이다.

EWM 모델의 예측 성능은 예측일과 2일 전의 기상 패턴 차이의 영향을 크게 받는다고 추측 가능하다. 예를 들어, 예측일과 2일 전의 날씨가 모두 맑아서 비슷할 경우 시계열 모델의 예측 성능은 좋을 것이고, 딥러닝 모델은 다양한 예보 변수 변동성의 영향으로 상대적으로 성능이 저하될 수 있다. 반면, 날씨가 극변하는 경우 기상 예보 오차가 크지 않다면 딥러닝 모델의 성능이 시계열 모델에 비해 성능이 좋을 것이다. 예측일 기준 2일전과 예측일의 날씨코드를 이용해 (맑음, 흐림)과 같은 방식으로 날씨들을 16가지 기상 패턴으로 분류하고, 각각의 경우에 대한 모델별 성능을 계산한 결과는 [그림 1]와 같다.



[그림 1] 기상 패턴 별 모델 성능

딥러닝 모델은 EWM 모델에 비해 성능 편차가 크지 않기 때문에 weather 모델은 제외하고 표기하였다. 모델별로 기상 조건에 따른 성능이 차이가 나는 것을 볼 수 있고, 특히 EWM 모델의 경우 기상 조건에 따라 다른 딥러닝 모델에 비해 성능이 더 좋거나 나쁠 수 있다.

II.2.2. 기상 패턴 기반 예측 모델 선택 방법

예측 시점의 조건에 따라 모델의 성능이 상이하기 때문에 예측일과 기상 패턴이 가장 비슷한 날을 찾고, 해당 일의 3가지 모델의 성능을 비교한 후 가장 성능이 좋았던 모델을 선택하는 로직을 적용할 수 있다. 이 때, 가장 기상 조건이 비슷한 날은 유클리드 거리(Euclidean Distance) 기반으로 계산되며, 거리가 가장 최소화되는 날로 정해진다. 그 거리는 식 (3)과 같이 계산된다.

$$dist = \sum_{t=6}^{19} \left\{ \left(W_t^{d_{target}} - W_t^{d_{candidate}} \right)^2 + \left(W_t^{d_{target}-2} - W_t^{d_{candidate}-2} \right)^2 \right\} \quad (3)$$

W_t 는 해당일의 시간대별 날씨코드이고, 최대 2주의 과거 날짜를 대상으로 비교한다. d_{target} 은 예측일($d+1$), $d_{candidate}$ 은 이전 날짜로, $d_{candidate} \in \{d-15, \dots, d-2\}$ 이다. 태양광이 일반적으로 발전하는 시간대가 주간 시간대인 6~19시이므로, t 의 값은 6~19시로 한정하여 계산하였다. 2일 간격의 두 기상 패턴을 동시에 비교하는 이유는 하루 전 예측의 특성 상 EWM 모델의 성능이 예측일과 2일 전의 기상 패턴 차이에 절대적으로 의존하기 때문이다.

최종 모델의 예측값은 기상 조건이 가장 비슷한 날에 모델별로 해당일의 발전량 MAE 비교 후 가장 오차가 낮은 모델의 예측값을 선택한다.

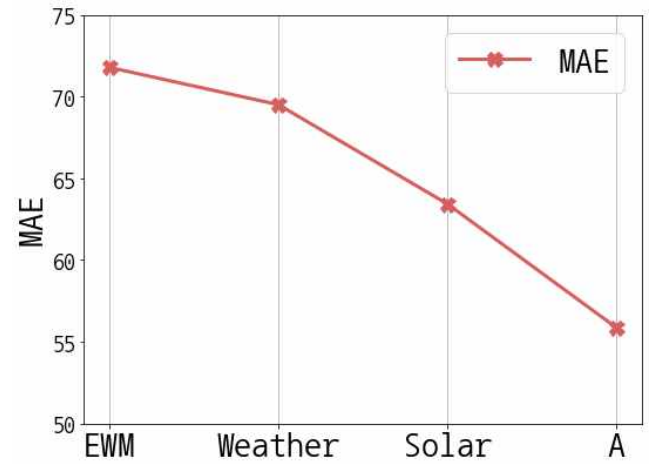
II.3. 실험 결과

제안한 로직의 검증을 위해 2022년 3월부터 2024년 4월까지 한 시간 단위로 계측된 발전량 데이터를

사용하였다. 기상 데이터는 기상청에서 제공하는 한 시간 단위의 예보 데이터를 사용하였다. EWM 모델은 예측 전날부터 최대 30일 전까지의 발전량 데이터를 사용하여 예측값을 생성하였고, 나머지 딥러닝 모델의 훈련은 예측 전날부터 축적된 데이터 중 최대 1년 전까지의 데이터를 사용하여 모델을 구축하였다.

실험은 2023년 11월부터 2024년 4월까지의 발전소의 모델별 예측값을 생성한 후, 3개의 모델과 본 논문에서 제시한 기상 패턴이 가장 비슷한 날의 최종 모델을 선택했을 때(A)의 평균 MAE로 성능을 평가하였다.

각 모델과 (A)의 MAE를 계산한 결과는 [그림 2]와 같다. (A)는 약 55.8kWh로 다른 모델 및 로직에 비해 성능이 우수함을 확인할 수 있다. MAE의 표준편차는 EWM이 226.19, weather 모델이 212.94, solar 모델이 212.19, (A)가 191.73으로, 예측 오차의 변동성 측면에서도 논문에서 제안한 방식이 우수하다.



[그림 2] EWM, weather, solar, (A)의 MAE

III. 결론

본 논문에서는 예측일과 기상 패턴이 가장 비슷한 날의 서로 다른 3가지 예측 모델의 오차를 참고하여 최종 모델을 선택하는 로직을 제안했다. 예측 모델 별로 입력 변수와 메커니즘이 상이하여 기상 환경에 따라 성능 차가 발생하는데, 이 부분을 고려하여 최종 예측값을 선택하는 것이 예측 오차 및 변동성 측면에서 효과적임을 검증하였다. 이외에도 기상 패턴을 다른 기상 변수들을 활용하거나, 모델 자체의 성능을 올리는 등의 방법을 통해 추가 개선할 수 있다.

참 고 문 헌

- [1] 손혜숙, 김석연, 장운, "기상 예보를 활용한 LSTM 기반 24 시간 태양광 발전량 예측 모델", 정보과학회 컴퓨팅의 실제 논문지, 26(10), pp. 435-441, 2020.
- [2] 김백천, 정승환, 김민석, 김종근, 김성신, "계절별 기상조건에 기반한 태양광 발전량 예측에 관한 연구", 한국지능시스템학회 논문지, 31(2), pp. 102-108, 2021.
- [3] P. Li, K. Zhou and S. Yang, "Photovoltaic Power Forecasting: Models and Methods", 2nd IEEE Conf. Energy Internet Energy Syst. Integr. EI2 2018 - Proc., pp. 1-6, 2018.