

# 전이 학습 기반의 랜덤 포레스트를 이용한 일사량 예측 기법

소다영<sup>1</sup>, 김의년<sup>1</sup>, 한태규<sup>1</sup>, 하휘명<sup>2</sup>, 문지훈<sup>1</sup>

<sup>1</sup>순천향대학교 빅데이터공학과

<sup>2</sup>LG 에너지솔루션

<sup>1</sup>{dayeong, eui20n, gksxorb159, jmoon22}@sch.ac.kr

<sup>2</sup>hwmhkr@lgensol.com

## A Solar Irradiance Forecasting Method Using Transfer Learning-Based Random Forests

Dayeong So<sup>1</sup>, Euinyun Kim<sup>1</sup>, Taegyung Han<sup>1</sup>, Hwimyeong Ha<sup>2</sup>, and Jihoon Moon<sup>1</sup>

<sup>1</sup>Dept. of Big Data Engineering, Soonchunhyang University

<sup>2</sup>LG Energy Solution, Ltd.

### 요약

일사량 정보가 없는 지역에 태양광 발전 시스템을 설치한다고 가정할 때, 수집된 일사량 데이터가 충분하지 않아 고도화된 인공지능 모델을 구성하기가 어려운데 이럴 때는 어떻게 정확한 일사량 예측을 할 수 있을까요? 본 논문은 일사량 정보가 부족한 지역에서 초기 태양광 발전 시스템의 효율적인 운영을 위해 전이학습 기반의 일사량 예측 기법을 제안한다. 먼저, 충분한 일사량 데이터가 수집된 여러 지역의 일사량과 기온, 습도, 풍속, 풍향과 같은 동네예보에서 제공하는 기상 정보를 수집하여 학습 데이터 셋을 구성한다. 다음으로 랜덤 포레스트의 빠른 구현이 가능한 Ranger 를 이용하여 학습 데이터 셋과 초기 태양광 발전 시스템의 설치 지역에 관한 하루치 기상 정보를 함께 학습하여 랜덤 포레스트 모델을 구성하고 다음 날의 하루치 일사량 예측을 수행한다. 제안한 방법론은 슬라이딩 윈도우 방식으로 모의실험을 통해 지속성(Persistence) 및 하루치로 학습한 랜덤 포레스트 모델들보다 우수한 예측 성능을 도출하여, 초기 태양광 발전 시스템에서 만족스러운 일사량 예측 성능을 도출할 수 있음을 확인하였다.

### 1. 서론

최근 화석연료 사용으로 인한 온실가스 배출량을 줄이기 위해 신재생 에너지와 이를 적극적으로 활용하는 스마트 그리드 기술의 관심이 커지고 있다[1-3]. 특히 신재생 에너지 중에서도 지속 가능한 에너지인 태양광은 이산화탄소를 거의 발생하지 않으며 석유 의존도를 낮추는 데 이바지할 수 있어 지구 온난화와 에너지 위기를 대처할 수 있는 핵심 기술로 많은 주목을 받고 있다[1]. 태양광 발전 시스템의 다양한 기상 요인들은 전력 생산에 영향을 주며, 태양광 발전 시스템은 미래의 일사량 정보를 기반으로 향후 생산될 전력을 통해 에너지 운영 계획을 수립하기 때문에 일사량 예측은 태양광 발전 시스템에서 매우 중요한 역할을 한다[2].

하지만 태양광 발전 시스템에서 전력 생산량에 영향을 미치는 온도, 습도, 풍속, 풍향 등과 같은 기상 정보는 기상청의 동네예보 시스템을 통해 쉽게 수집 및 이용할 수 있지만, 가장 큰 영향을 미치는 미래 일사량 정보는 제공하지 않는다. 따라서 최적의 태양광 발전 시스템을 운영하기 위해서는 기상청의 동네예보를 통해 정확한 일사량 모델을 구성하여 미래 일사량 정보를 예측하는 것이 중요하다. 그리하여 기상청 동네예보에서 제공하는 기상 요인들과 일사량과의 관계를 효과적으로 학습하기

위해 기계학습 기반의 단기 일사량 예측 모델을 구성한 사례가 보고되었으며, 대다수의 모델은 예측 대상에서 충분한 기간 동안 수집된 데이터 셋을 이용하였다.

최근 태양광 시장이 크게 성장함에 따라 일사량 정보가 수집되지 않은 지역에도 태양광 발전 시스템을 구축하는 사례가 증가하고 있다. 하지만 해당 지역에는 기계학습을 적용하기에는 학습 데이터가 부족하므로 만족스러운 일사량 예측 성능을 기대하기가 매우 어렵다. 전이 학습(Transfer Learning)은 방대한 양의 데이터로 이미 학습된 인공지능을 유사하지만 다른 분야에 적용하는 방법이다[3]. 본 논문은 일사량 정보가 부족한 지역을 대상으로 정확한 단기 일사량 예측을 위해 전이 학습 기반의 랜덤 포레스트를 이용한 다단계 일사량 예측 모델을 제안하며, 태양광 발전 시스템에서 수집된 일사량 정보가 하루치만 존재하더라도 만족스러운 예측 성능을 도출하는 것을 목표로 둔다.

### II. 본론

본 논문은 먼저 대전을 일사량 정보가 없는 초기 태양광 발전 시스템의 설치 지역으로 선정하여, 서울, 부산, 대구, 인천, 광주에서는 충분한 일사량 데이터가 수집되었다고 가정하고 실험을 진행하였다. 이를 위해 서울, 부산, 대구,

Table 1. 예측 모델 학습을 위한 입력 변수 구성

No.	Variable Name	Description
1	Month	Month of the year
2	Day	Day of the month
3	Hour	Hour of the day
4	Temp	Temperature
5	Humi	Humidity
6	WS	Wind speed
7	WD	Wind direction

인천, 광주 지역의 일사량과 기상청 동네예보를 통해 미래 기상 정보를 확인할 수 있는 기온, 습도, 풍속, 풍향 데이터를 2016 년 1 월 1 일 오전 8 시부터 2019 년 12 월 31 일 오후 6 시까지 수집하였으며, 제한한 예측 모델의 효용성을 입증하기 위해 대전의 일사량을 포함한 여러 기상 정보들을 2020 년 1 월 1 일 오전 8 시부터 12 월 31 일 오후 6 시까지 1 시간 단위로 수집하였다.

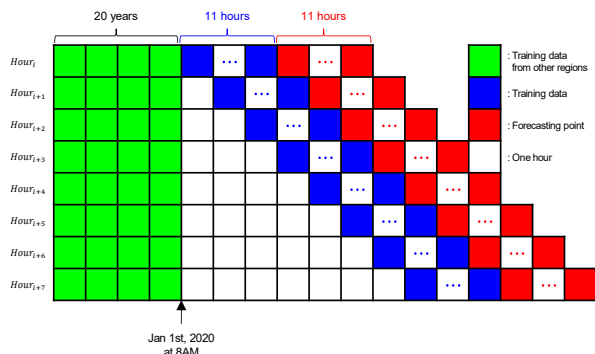


Figure 1. 전이 학습을 이용한 다단계 일사량 예측 방식

전이 학습 모델링을 위해, Table 1 과 같이 7 가지의 입력 변수들을 구성하고 다섯 군데 지역(서울, 부산, 대구, 인천, 광주)의 4 년치 데이터 셋과 예측 대상 지역의 1 년치 데이터 셋을 결합하였다. 또한, 태양광 발전 시스템이 하루만 운영되었으며, 일사량 데이터도 하루치만 수집되었다는 것을 전제로 모의실험을 진행하기 위해 Figure 1 과 같이 학습 데이터를 구성하였다. 예를 들어 2020 년 3 월 1 일 오전 8 시부터 오후 6 시까지 일사량을 예측한다면, 서울, 부산, 대구, 인천, 광주의 2016 년부터 2019 년까지의 데이터와 대전의 2020 년 2 월 28 일 오전 8 시부터 오후 6 시까지의 데이터를 이용하였다.

랜덤 포레스트는 정확성, 단순성, 유연성 등 여러 장점으로 인해 다양한 분야에서 주로 사용되는 기계학습 기법으로, R 에서 일반적으로 랜덤 포레스트 모델을 구성하기 위해 randomForest 패키지가 널리 사용된다. 그러나 여러 차원의 데이터 사용에 최적화되지 않았으며, 모델을 구현하는 속도가 느리다는 단점이 있다. 따라서 본 연구는 랜덤 포레스트의 빠른 구현인 레인저(RANGER; RANdom Forest GEnerator) 패키지[4]를 사용하여 랜덤 포레스트 모델을 구성하였다. 또한, 랜덤 포레스트의 초매개변수인 나무의 수는 Oshiro 등[5]의 연구에서 제안한 값인 128 을 사용하였다.

Table 2 는 하루치만 학습한 랜덤 포레스트(Baseline Model)와 제안한 모델(Ours)과의 성능을 RMSE (Root Mean Square Error)와 MAE (Mean Absolute Error)

Table 2. 예측 모델 간의 성능 비교 (단위: MJ/m<sup>2</sup>)

Steps	Baseline Model		Ours	
	RMSE	MAE	RMSE	MAE
1	0.689	0.559	0.518	0.393
2	0.793	0.643	0.527	0.399
3	0.847	0.683	0.528	0.402
4	0.878	0.706	0.529	0.402
5	0.872	0.700	0.530	0.402
6	0.856	0.685	0.531	0.403
7	0.834	0.663	0.531	0.404
8	0.815	0.642	0.531	0.403
9	0.802	0.629	0.532	0.404
10	0.806	0.635	0.532	0.404
11	0.831	0.664	0.533	0.404
Avg.	0.820	0.655	0.529	0.402

관점에서 비교 분석한 결과를 제시한 것이다. 제안한 모델은 다른 지역의 충분한 데이터를 통해 기상 요인과 일사량과의 관계를 효과적으로 학습할 수 있어 더욱 정확한 예측 성능을 도출하였다. 또한, 11 번째 시점을 기준으로 제안한 모델은 현재 시간의 일사량 값을 다음 날 같은 시간의 일사량 값과 동일하다고 가정한 지속성 모델(RMSE: 0.839, MAE: 0.565)과 비교하였을 때, 더욱 우수한 성능을 도출하여 초기 태양광 발전 시스템에서 제안한 모델의 적용 가능성을 확인할 수 있었다.

### III. 결 론

본 논문은 초기 태양광 발전 시스템에서 강건한 일사량 예측을 위한 전이 학습 기반의 랜덤 포레스트 모델을 개발하였다. 제안한 모델은 예측 대상에서 하루치의 일사량만 수집되었다 하더라도 다른 지역의 충분한 일사량 데이터를 활용하여 일사량과 여러 기상 요인과의 관계를 효과적으로 학습할 수 있었다. 향후 서울, 부산, 대구, 인천, 광주 등 다양한 지역을 대상으로 본 연구의 범용성을 확인할 예정이다.

### Acknowledgement

본 논문은 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구 결과로 수행되었음(2021-0-01399).

### 참 고 문 헌

- [1] Y. Kwon, A. Kwasinski, and A. Kwasinski, "Solar Irradiance Forecast Using Naïve Bayes Classifier Based on Publicly Available Weather Forecasting Variables," *Energies*, Vol. 12, No. 8, p. 1529, 2019.
- [2] D. Yang, N. Yeo and P. Mah, "A Dynamic Piecewise Prediction Model of Solar Insolation for Efficient Photovoltaic Systems," *KIIE Transactions on Computing Practices*, Vol. 23, No. 11, pp. 632-640, 2017.
- [3] J. Moon, J. Kim, P. Kang, and E. Hwang, "Solving the Cold-Start Problem in Short-Term Load Forecasting Using Tree-Based Methods," *Energies*, Vol. 13, No. 4, p. 886, 2020.
- [4] D. Yang, N. Yeo and P. Mah, "A Dynamic Piecewise Prediction Model of Solar Insolation for Efficient Photovoltaic Systems," *KIIE Transactions on Computing Practices*, Vol. 23, No. 11, pp. 632-640, 2017.
- [5] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How Many Trees in a Random Forest?" In *Proc. of the International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 154-168, 2012.